RESEARCH CENTER

FIELD

Activity Report 2012

Section New Results

Edition: 2013-04-24

BYMOORE Exploratory Action

4. New Results

4.1. BenchNN: On the Broad Potential Application Scope of Hardware Neural Network Accelerators

The emergence of high-performance applications like Recognition, Mining, and Synthesis (RMS) suggest that the potential application scope of a hardware neural network accelerator would be broad. We have highlighted that a hardware neural network accelerator is indeed compatible with many of the emerging high-performance workloads, currently accepted as benchmarks for high-performance micro-architectures. For that purpose, we develop and evaluate software neural network implementations of 5 (out of 12) RMS applications from the PARSEC Benchmark Suite. Our results show that neural network implementations can achieve competitive results, with respect to application-specific quality metrics, on these 5 RMS applications.

4.2. A Defect-Tolerant Accelerator for Emerging High-Performance Applications

Due to the evolution of technology constraints, especially energy constraints which may lead to heterogeneous multicores, and the increasing number of defects, the design of defect-tolerant accelerators for heterogeneous multi-cores may become a major micro-architecture research issue. Most custom circuits are highly defect sensitive, a single transistor can wreck such circuits. On the contrary, artificial neural networks (ANNs) are inherently error tolerant algorithms. And the emergence of high-performance applications implementing recognition and mining tasks, for which competitive ANN-based algorithms exist, drastically expands the potential application scope of a hardware ANN accelerator. However, while the error tolerance of ANN algorithms is well documented, there are few in-depth attempts at demonstrating that an actual hardware ANN would be tolerant to faulty transistors. Most fault models are abstract and cannot demonstrate that the error tolerance of ANN algorithms can be translated into the defect tolerance of hardware ANN accelerators. In this article, we introduce a hardware ANN geared towards defect tolerance and energy efficiency, by spatially expanding the ANN. In order to precisely assess the defect tolerance capability of this hardware ANN, we introduce defects at the level of transistors, and then assess the impact of such defects on the hardware ANN functional behavior. We empirically show that the conceptual error tolerance of neural networks does translate into the defect tolerance of hardware neural networks, paving the way for their introduction in heterogeneous multi-cores as intrinsically defect-tolerant and energy-efficient accelerators.

4.3. Design of a Hardware Spiking Neural Network

Hardware Spiking Neurons Design: Analog or Digital? Neurons can be implemented either as analog or digital components. While the respective advantages of each approach are well known, i.e., digital designs are more simple but analog neurons are more energy efficient, there exists no clear and precise quantitative comparison of both designs. In this paper, we compare the digital and analog implementations of the same Leaky Integrate-and-Fire neuron model at the same technology node (CMOS 65 nm) with the same level of performance (SNR and maximum spiking rate), in terms of area and energy. We show that the analog implementation requires 5 times less area, and consumes 20 times less energy than the digital design. As a result, the analog neuron, in spite of its greater design complexity, is a serious contender for future large-scale silicon neural systems.

Configurable Conduction Delay Circuits for High Spiking Rates. The conduction delay in neural systems has been proven to play an important role in processing neural information. In hardware spiking neural networks (SNN), emulating conduction delays consists of intercepting and buffering spikes for a certain amount of time during their transfer. The complexity of the conduction delay implementation increases with high spiking rates; it implies (1) storing a large number of spikes into memory cells and (2) conserving the required time resolution while processing the delays. As a result, the circuit size becomes very large and difficult to integrate into large scale SNN systems. In this paper, we highlight the trade-offs of an efficient digital delay circuit design supporting high neuron firing rates. The key issue resides in conserving spikes and spike timings while limiting storage requirements. We present a digital implementation of a configurable delay circuit supporting spiking rates of up to 1Meps (Mega events per second) and a delay range going from 1μs to 50ms with a time resolution less than 5% of the configured delay time. Synthesis results show that, using the CMOS 65nm technology, the required silicon area is 1600um2.

4.4. 3D-Stacked Implementation of Neural Networks

In order to cope with increasingly stringent power and variability constraints, architects need to investigate alternative paradigms. Neuromorphic architectures are increasingly considered (especially spike-based neurons) because of their inherent robustness and their energy efficiency. Yet, they have two limitations: the massive parallelism among neurons is hampered by 2D planar circuits, and the most cost-effective hardware neurons are analog implementations that require large capacitors, We show that 3D stacking with Through-Silicon-Vias applied to neuromorphic architectures can solve both issues: not only by providing massive parallelism between layers, but also by turning the parasitic capacitances of TSVs into useful capacitive storage.

4.5. Iterative Optimization for the Data Center (Alchemy-related research)

This result corresponds to research started within Alchemy, and it is less related to the objectives of ByMoore itself.

Iterative optimization is a simple but powerful approach that searches for the best possible combination of compiler optimizations for a given workload. However, each program, if not each data set, potentially favors a different combination. As a result, iterative optimization is plagued by several practical issues that prevent it from being widely used in practice: a large number of runs are required for finding the best combination; the process is inherently data set sensitive; and the exploration process incurs significant overhead that needs to be compensated for by performance benefits. Therefore, while iterative optimization has been shown to have significant performance potential, it is seldomly used in production compilers.

We propose [5] Iterative Optimization for the Data Center (IODC): we show that servers and data centers offer a context in which all of the above hurdles can be overcome. The basic idea is to spawn different combinations across workers and recollect performance statistics at the master, which then evolves to the optimum combination of compiler optimizations. IODC carefully manages costs and benefits, and is transparent to the end user.

We evaluate IODC using both MapReduce and throughput server applications. In order to reflect the large number of users interacting with the system, we gather a very large collection of data sets (at least 1000 and up to several million unique data sets per program), for a total storage of 10.7TB, and 568 days of CPU time. We report an average performance improvement of $1.48\times$, and up to $2.08\times$, for the MapReduce applications, and $1.14\times$, and up to $1.39\times$, for the throughput server applications.

4.6. Statistical Performance Comparisons of Computers (Alchemy-related research)

This result corresponds to research started within Alchemy, and it is less related to the objectives of ByMoore itself.

As a fundamental task in computer architecture research, performance comparison has been continuously hampered by the variability of computer performance. In traditional performance comparisons, the impact of performance variability is usually ignored (i.e., the means of performance measurements are compared regardless of the variability), or in the few cases where it is factored in using parametric confidence techniques, the confidence is either erroneously computed based on the distribution of performance measurements, instead of the distribution of sample mean of performance measurem ents, or too few measurements are considered for the distribution of sample mean to be normal. We first illustrate how such erroneous practices can lead to incorrect comparisons.

Then, we propose [4] a non-parametric Hierarchical Performance Testing (HPT) framework for performance comparison, which is significantly more practical than standard p arametric confidence tests because it does not require to collect a large number of measurements in order to achieve a normal distribution of the sample mean. This HPT framework has been implemented as an open-source software.

POPIX Exploratory Action

6. New Results

6.1. Mixture of mixed effects models

Participants: Cyprien Mbogning, Marc Lavielle.

We have proposed a new methodology for maximum likelihood estimation in mixtures of non linear mixed effects models (NLMEM). The article *Inference in mixtures of non-linear mixed effects models* was submitted in 2012. Such mixtures of models include mixtures of distributions, mixtures of structural models and mixtures of residual error models. Since the individual parameters inside the NLMEM are not observed, we propose to combine the EM algorithm usually used for mixtures models when the mixture structure concerns an observed variable, with the Stochastic Approximation EM (SAEM) algorithm, which is known to be suitable for maximum likelihood estimation in NLMEM and also has nice theoretical properties. The main advantage of this hybrid procedure is to avoid a simulation step of unknown group labels required by a "full" version of SAEM. The resulting MSAEM (Mixture SAEM) algorithm is now implemented in the MONOLIX software. Several criteria for classification of subjects and estimation of individual parameters were also proposed. Numerical experiments on simulated data have shown that MSAEM performs well in a general framework of mixtures of NLMEM. Indeed, MSAEM provides an estimator close to the maximum likelihood estimator in very few iterations and is robust with regard to initialization. An application to pharmacokinetic (PK) data demonstrates the potential of the method for practical applications.

6.2. Between-subject and within-subject model mixtures for classifying HIV treatment response

Participants: Cyprien Mbogning, Kevin Bleakley, Marc Lavielle.

We have proposed a method for classifying individuals into clinically-relevant population subgroups [5]. This is achieved by treating "subgroup" as a categorical covariate whose value is unknown for each individual, and predicting its value using mixtures of models that represent "typical" longitudinal data from each subgroup. Under a nonlinear mixed effects model framework, two types of model mixtures were developed:

- *Between-Subject Model Mixtures* (BSMM) assume that each individual's longitudinal data follows one of M "base" models, but we do not necessarily know *a priori* which one. Individual i thus has a label $z_i = m \in \{1, ..., M\}$ referring to the model that is supposed to have generated it. We have shown how to extract *a posteriori* estimates of the probability that each individual was generated by each of the base models; this can be used to predict which type of patient we have: non-responder, responder or rebounder.
- *Within-Subject Model Mixtures* (WSMM) make the hypothesis that the model mixture occurs *within each individual*. In the HIV example, this means that we consider that each patient is partially a non-responder, partially a responder and partially a rebounder. This is perhaps more biologically plausible than BSMMs in the sense that each individual's response may be due to their own particular combination of virus strains, cell populations, etc. Within the NLMEM framework, this means including individual "model proportion" parameters into the model and having to estimate them along with the other parameters of the NLMEM. It turns out that this does not require any mathematical extensions to a typical NLMEM. But we can use the estimated proportions to help categorize patients, especially those who do not naturally fall into one of the three "typical" categories.

An application to longitudinal viral load data for HIV-positive patients were used to predict whether they are responding – completely, partially or not at all – to a new drug treatment.

6.3. Joint modeling of longitudinal and repeated time-to-event data

Participants: Cyprien Mbogning, Kevin Bleakley, Marc Lavielle.

We have proposed a nonlinear mixed-effects framework to jointly model longitudinal and repeated time-to-event data. The article *Joint modeling of longitudinal and repeated time-to-event data with maximum likelihood estimation via the SAEM algorithm* was submitted in 2012. A parametric nonlinear mixed-effects model is used for the longitudinal observations and a parametric mixed-effects hazard model for repeated event times. We have shown the importance for parameter estimation of properly calculating the conditional density of the observations (given the individual parameters) in the presence of interval and/or right censoring. Parameters are estimated by maximizing the exact joint likelihood with the Stochastic Approximation Expectation-Maximization algorithm.

We have illustrated the use of these modeling methods in two real data examples: patient survival in primary biliary cirrhosis, and repeated epileptic seizure count data from a clinical trial.

This workflow for joint models is now implemented in the MONOLIX software

6.4. A new Bayesian Information Criteria for mixed-effects models

Participants: Maud Delattre, Marie-Anne Poursat, Marc Lavielle.

The Bayesian Information Criterion (BIC) is widely used for variable selection in mixed effects models. However, its expression is unclear in typical situations of mixed effects models, where simple definition of the sample size is not meaningful. Yet, in the mixed effects model literature, the BIC penalty usually involves the total number of observations $\log n_{\rm tot}$. From a practical point of view, the $\log n_{\rm tot}$ penalty is implemented in the R package nlme and in the SPSS procedure MIXED while the $\log N$ penalty, where N is the number of subjects, is used in MONOLIX, saemix or in the SAS proc NLMIXED.

We have derived an appropriate BIC expression that is consistent with the random effect structure of the mixed effects model [7]. We have illustrated the behavior of the proposed criterion through a simulation study. The use of this new version of BIC is recommended as an alternative to various existing BIC versions that are implemented in available software.

6.5. Inference in mixed hidden Markov models

Participants: Maud Delattre, Marc Lavielle.

Mixed hidden Markov models have been recently defined in the literature as an extension of hidden Markov models for dealing with population studies. The notion of mixed hidden Markov models is particularly relevant for modeling longitudinal data collected during clinical trials, especially when distinct disease stages can be considered. However, parameter estimation in such models is complex, especially due to their highly nonlinear structure and the presence of unobserved states. Moreover, existing inference algorithms are extremely time consuming when the model includes several random effects.

We have proposed new inference procedures for estimating population parameters, individual parameters and sequences of hidden states in mixed hidden Markov models [1]. The main contribution consists of a specific version of the stochastic approximation EM algorithm coupled with the Baum-Welch algorithm for estimating population parameters. The properties of this algorithm were investigated via a Monte-Carlo simulation study.

An application of mixed hidden Markov models to the description of daily seizure counts in epileptic patients was then considered. We proposed to describe exposure-response relationship of gabapentin in epileptic patients using MHMM approach. Longitudinal seizure frequency data from six clinical studies were available for the analysis. The model describes daily seizure frequencies to be governed by an unobserved, yet present, underlying disease dynamics, defined by states of high or low epileptic activity. Individual day-to-day states are dependent exhibiting their own dynamics with patients transitioning between disease states, according to a set of transition probabilities. MHMM estimates both unobserved disease dynamics and daily seizure frequencies in all disease states. Novel drug action modes are achievable: drug may influence both seizure frequencies and transition probabilities. The model showed that gabapentin significantly reduced seizure frequencies in both disease states, without altering disease dynamics. Novel methodology offers additional insights into understanding epilepsy time course, gabapentin mode of action and provides a tool for realistic clinical trial simulations.

6.6. Inference in mixed-effects diffusion models

Participants: Maud Delattre, Marc Lavielle.

The structure of mixed effects models allows a suitable consideration of the whole variability characterizing such data, which is usually split into some intra-individual variability - i.e., the variability occurring within the dynamics of each individual - and some between-subjects variability. In a mixed-effects model, the same structural model is used for describing each individual sequence of observations, but the parameters of this model vary randomly among the individuals, which allows a correct account of the differences between subjects. In a mixed-effects diffusion model, the description of each individual series of observations is based on stochastic differential equations (SDEs). Diffusion is known to be a relevant tool for describing random variability in dynamical systems, and is widely used in applications in many domains.

Although many methods are available for the inference in classical fixed-effects diffusion models, there is still a need for a general, fast and easy to implement method for the inference in mixed-effects diffusion models. Indeed, except in very specific classes of mixed-effects diffusion models, the likelihood of the observations does not have any closed-form expression, making maximum likelihood estimation of the model parameters an intricate issue. The difficulty is twofold for computing the observed likelihood since it involves the transition densities of the underlying individual diffusion processes and integrals over the unobserved individual parameters that can rarely be computed in a closed form. Specific versions of the SAEM algorithm have already been proposed for estimating the population parameters in mixed-effects diffusion models (using for instance an Euler-Maruyama approximation of the individual processes or some particle Markov Chain Monte-Carlo methods). In these two versions of SAEM however, simulation of both the random individual parameters and the individual latent processes is required at simulation step, which is computationally cumbersome.

We have proposed a new inference methodology for mixed-effects diffusion models which consists in coupling the SAEM algorithm with the extended Kalman filter for estimating the population parameters. The relevant article has been submitted in 2012. In this new version of the SAEM algorithm, we only need to simulate the individual parameters at each iteration. We also provide tools for estimating the individual parameters and the individual diffusion trajectories.

6.7. Random threshold for linear model selection

Participant: Marc Lavielle.

We have in a previous work introduced a random thresholding method to select the significant, or non-null, mean terms from a collection of independent random variables, and applied it to the problem of recovering the significant coefficients in nonordered model selection.

We have improved this method by introducing a simple modification which removes the dependency of the proposed estimator on a window parameter while maintaining its asymptotic properties [4]. A simulation study suggests that both procedures compare favorably to standard thresholding approaches, such as multiple testing or model-based clustering, in terms of the binary classification risk. An application to the problem of activation detection on functional magnetic resonance imaging (fMRI) data was used to illustrate the performance of the proposed method.

ABSTRACTION Project-Team

6. New Results

6.1. Analysis of Biological Pathways

We have improved our framework to design and analyze biological networks. This framework focused on protein-protein interaction networks described as graph rewriting systems. Such networks can be used to model some signaling pathways that control the cell cycle. The task is made difficult due to the combinatorial blow up in the number of reachable species (*i.e.*, non-isomorphic connected components of proteins).

6.1.1. Semantics

Participants: Jonathan Hayman, Tobias Heindel [CEA-List].

Domain-specific rule-based languages can be understood intuitively as transforming graph-like structures, but due to their expressivity these are difficult to model in 'traditional' graph rewriting frameworks.

In [21], we introduce pattern graphs and closed morphisms as a more abstract graph-like model and show how Kappa can be encoded in them by connecting its single-pushout semantics to that for Kappa. This level of abstraction elucidates the earlier single-pushout result for Kappa, teasing apart the proof and guiding the way to richer languages, for example the introduction of compartments within cells.

6.1.2. Semantics and causality

Participants: Vincent Danos [University of Edinburgh], Jérôme Feret, Walter Fontana [Harvard Medical School], Russ Harmer [Paris VII], Jonathan Hayman, Jean Krivine [Paris VII], Chris Thompson-Walsh [University of Cambridge], Glynn Winskel [University of Cambridge].

In [20], we introduce a novel way of constructing concise causal histories (pathways) to represent how specified structures are formed during simulation of systems represented by rulebased models. This is founded on a new, clean, graph-based semantics introduced in the first part of this paper for Kappa, a rule-based modelling language that has emerged as a natural description of protein-protein interactions in molecular biology. The semantics is capable of capturing the whole of Kappa, including subtle side-effects on deletion of structure, and its structured presentation provides the basis for the translation of techniques to other models. In particular, we give a notion of trajectory compression, which restricts a trace culminating in the production of a given structure to the actions necessary for the structure to occur. This is central to the reconstruction of biochemical pathways due to the failure of traditional techniques to provide adequately concise causal histories, and we expect it to be applicable in a range of other modelling situations.

6.1.3. Case study: Combinatorial drift in yeast model

Participants: Vincent Danos [University of Edinburgh], Eric Deeds [University of Kansas], Jérôme Feret, Walter Fontana [Harvard Medical School], Russ Harmer [Paris VII], Jean Krivine [Paris VII].

The assembly of molecular machines and transient signaling complexes does not typically occur under circumstances in which the appropriate proteins are isolated from all others present in the cell. Rather, assembly must proceed in the context of large-scale protein-protein interaction (PPI) networks that are characterized both by conflict and combinatorial complexity. Conflict refers to the fact that protein interfaces can often bind many different partners in a mutually exclusive way, while combinatorial complexity refers to the explosion in the number of distinct complexes that can be formed by a network of binding possibilities.

In [9], we use computational models so as to explore the consequences of these characteristics for the global dynamics of a PPI network based on highly curated yeast two-hybrid data. The limited molecular context represented in this data-type translates formally into an assumption of independent binding sites for each protein. The challenge of avoiding the explicit enumeration of the astronomically many possibilities for complex formation is met by a rule-based approach to kinetic modeling. Despite imposing global biophysical constraints, we find that initially identical simulations rapidly diverge in the space of molecular possibilities, eventually sampling disjoint sets of large complexes. We refer to this phenomenon as “compositional drift". Since interaction data in PPI networks lack detailed information about geometric and biological constraints, our study does not represent a quantitative description of cellular dynamics. Rather, our work brings to light a fundamental problem (the control of compositional drift) that must be solved by mechanisms of assembly in the context of large networks. In cases where drift is not (or cannot be) completely controlled by the cell, this phenomenon could constitute a novel source of phenotypic heterogeneity in cell populations.

6.1.4. Automatic Reduction of Stochastic Semantics

Participants: Ferdinanda Camporesi, Jérôme Feret, Norman Ferns, Thomas Henzinger [Institute of Science and Technology, Austria], Heinz Koeppl [ETH Zürich], Tatjana Petrov [ETH Zürich].

Biology, Protein-protein interaction networks, Stochastic semantics, Verification.

We have proposed an abstract interpretation-based framework for reducing the state-space of stochastic semantics for protein-protein interaction networks. Our framework ensures that the trace distribution of the reduced system is the exact projection of the trace distribution of the concrete system. Moreover, when the abstraction is complete, if each state with the same abstraction is equiprobable at initial state, each state with the same abstraction is equiprobable at any time t.

In [10], we have formalized the model reduction framework for the stochastic semantics and we have established the relationships with the notions of lumpability, and bisimulation.

In [13], we have showed that the reduced models can be expressed in Kappa, and we have provided a procedure to do it.

6.2. Leakage Analysis

Participants: Matteo Zanioli [Correspondent], Pietro Ferrara [ETH, Zurich], Agostino Cortesi [Università Ca’ Foscari].

Abstract interpretation, Information leakage analysis, Object-oriented software, Static analysis.

In [28], we present SAILS, a new tool that combines SAMPLE, a generic static analyzer, and a sophisticated domain for leakage analysis. This tool does not require to modify the original language, since it works with mainstream languages like JAVA™, and it does not require any manual annotation. SAILS can combine the information leakage analysis with different heap abstractions, inferring information leakage over programs with complex data structures. SAILS has been applied to the analysis of the SecuriBench-micro suite. The experimental results underline the effectiveness of the analysis, since SAILS is in position to analyze several benchmarks in about 1 second without producing false alarms in more than 90% of the programs.

6.3. Termination

Participants: Patrick Cousot, Radhia Cousot.

Abstract interpretation, Computational induction, Induction, Proof, Static analysis, Semantic structural induction, Syntactic structural induction, Termination, Variant function, Verification.

In [17], we have introduced an abstract interpretation for termination.

Proof, verification and analysis methods for termination all rely on two induction principles: (1) a variant function or induction on data ensuring progress towards the end and (2) some form of induction on the program structure.

So far, no clear design principle did exist for termination as is the case for safety so that the existing approaches are scattered and largely not comparable with each other.

- For (1), we show that this design principle applies equally well to potential and definite termination. The trace-based termination collecting semantics is given a fixpoint definition. Its abstraction yields a fixpoint definition of the best variant function. By further abstraction of this best variant function, we derive the Floyd/Turing termination proof method as well as new static analysis methods to effectively compute approximations of this best variant function.
- For (2), we introduce a generalization of the syntactic notion of structural induction (as found in Hoare logic) into a semantic structural induction based on the new semantic concept of inductive trace cover covering execution traces by segments, a new basis for formulating program properties. Its abstractions allow for generalized recursive proof, verification and static analysis methods by induction on both program structure, control, and data. Examples of particular instances include Floyd's handling of loop cut-points as well as nested loops, Burstall's intermittent assertion total correctness proof method, and Podelski-Rybalchenko transition invariants.

6.4. Probabilistic Abstract Interpretation

Participants: Patrick Cousot, Michaël Monerau.

Abstract interpretation, Probabilistic systems, Static analysis.

Abstract interpretation has been widely used for verifying properties of computer systems. In [19], we present a way to extend this framework to the case of probabilistic systems.

The probabilistic abstraction framework that we propose allows us to systematically lift any classical analysis or verification method to the probabilistic setting by separating in the program semantics the probabilistic behavior from the (non-)deterministic behavior. This separation provides new insights for designing novel probabilistic static analyses and verification methods.

We define the concrete probabilistic semantics and propose different ways to abstract them. We provide examples illustrating the expressiveness and effectiveness of our approach.

6.5. Formal Verification by Abstract Interpretation

Participant: Patrick Cousot.

Abstract interpretation, Abstraction, Aerospace, Certification, Cyber-physical system, Formal Method, Mission-critical system, Runtime error, Safety-critical system, Scalability, Soundness, Static Analysis, Validation, Verification.

Abstract interpretation is a theory of abstraction and constructive approximation of the mathematical structures used in the formal description of programming languages and the inference or verification of undecidable program properties. Developed in the late seventies with Radhia Cousot, it has since then been considerably applied to many aspects of programming, from syntax, to semantics, and proof methods where abstractions are sound and complete but incomputable to fully automatic, sound but incomplete approximate abstractions to solve undecidable problems such as static analysis of infinite state software systems, contract inference, type inference, termination inference, model-checking, abstraction refinement, program transformation (including watermarking), combination of decision procedures, security, malware detection, etc.

This last decade, abstract interpretation has been very successful in program verification for mission- and safety-critical systems [12]. An example is ASTRÉE which is a static analyzer to verify the absence of runtime errors in structured, very large C programs with complex memory usages, and involving complex boolean as well as floating-point computations (which are handled precisely and safely by taking all possible rounding errors into account), but without recursion or dynamic memory allocation. Astrée targets embedded applications as found in earth transportation, nuclear energy, medical instrumentation, aeronautics and space flight, in particular synchronous control/command such as electric flight control or more recently asynchronous systems as found in the automotive industry. Astrée is industrialized by AbsInt Angewandte Informatik GmbH.

6.6. Static Analysis of Parallel Software

Participant: Antoine Miné.

Abstract interpretation, Embedded software, Parallel software, Rely/guarantee analysis, Run-time errors, Static analysis.

We present in [11] the theoretical foundation and the latest experimental evaluation of ASTRÉEA (5.3), a static analyzer prototype based on abstract interpretation to check for run-time errors in multi-threaded embedded critical C programs. Our method is based on a slightly modified non-parallel analysis that, when analyzing a thread, applies and enriches an abstract set of thread interferences. An iterator then re-analyzes each thread in turn until interferences stabilize. We prove the soundness of our method with respect to the sequential consistency semantics, but also with respect to a reasonable weakly consistent memory semantics. We also show how to take into account mutual exclusion and thread priorities through a partitioning over an abstraction of the scheduler state. This work is an extension of [54], complete with a full formalization and soundness proofs.

In [24], we express rely/guarantee methods in constructive form as an abstract interpretation of the interleaving trace semantics. We also restate the analysis presented in [11] as a further abstraction of rely/guarantee. This theoretical work brings a new understanding of the various causes of incompleteness and imprecision in our previous analysis, including the non-relational, input-insensitive, flow-insensitive, and history-insensitive treatment of interferences, and it opens the way to designing more precise analyses.

6.7. Static Analysis of Bit-Level Machine Integer and Floating-Point Operations

Participant: Antoine Miné.

Abstract interpretation, Embedded software, Numerical abstract domains, Run-time errors, Static analysis.

We present in [22] a few lightweight numeric abstract domains to analyze C programs that exploit the binary representation of numbers in computers, for instance to perform "compute-through-overflow" on machine integers, or to directly manipulate the exponent and mantissa of floating-point numbers. On integers, we propose an extension of intervals with a modular component, as well as a bitfield domain. On floating-point numbers, we propose a predicate domain to match, infer, and propagate selected expression patterns. These domains are simple, efficient, and extensible. We have included them into the ASTRÉE (5.2) and ASTRÉEA (5.3) static analyzers to supplement existing domains. Experimental results show that they can improve the analysis precision at a reasonable cost.

6.8. Inferring Sufficient Conditions with Backward Polyhedral Under-Approximations

Participant: Antoine Miné.

Abstract interpretation, Backward analysis, Numerical abstract domains, Static analysis, Sufficient condition inference, Under-approximations.

In [23], we discuss the automatic inference of sufficient pre-conditions by abstract interpretation and sketch the construction of an under-approximating backward analysis. We focus on numeric domains and propose transfer functions, including a lower widening, for polyhedra, without resorting to disjunctive completion nor complementation, while soundly handling non-determinism. A limited proof-of-concept prototype was designed to validate our approach. Planned applications include the derivation of sufficient conditions for a program to never step outside an envelope of safe states, or dually to force it to eventually fail.

6.9. A Constraint Solver Based on Abstract Domains

Participants: Marie Pelleau [University of Nantes, LINA], Antoine Miné, Charlotte Truchet [University of Nantes, LINA], Frédéric Benhamou [University of Nantes, LINA].

Abstract interpretation, Backward analysis, Numerical abstract domains, Static analysis, Sufficient condition inference, Under-approximations.

In [25], we apply techniques from abstract interpretation to constraint programming (which aims at solving hard combinatorial problems with a generic framework based on first-order logics). We highlight some links and differences between these fields: both compute fixpoints by iterations but employ different extrapolation and refinement strategies; moreover, consistencies in Constraint Programming can be mapped to non-relational abstract domains. We then use these correspondences to build an abstract constraint solver that leverages abstract interpretation techniques (such as relational domains) to go beyond classic solvers. We present encouraging experimental results obtained with our prototype implementation.

6.10. Automatic Inference of Necessary Preconditions

Participants: Patrick Cousot, Radhia Cousot, Manuel Fahndrich [Microsoft Research, Redmond, USA], Francesco Logozzo [Microsoft Research, Redmond, USA].

Abstract interpretation, Backward analysis, Static analysis, Necessary condition inference,

In [18], we consider the problem of automatic precondition inference for: (i) program verification; (ii) helping the annotation process of legacy code; and (iii) helping generating code contracts during code refactoring. We argue that the common notion of sufficient precondition inference (i.e., under which precondition is the program correct?) imposes too large a burden on call-sites, and hence is unfit for automatic program analysis. Therefore, we define the problem of necessary precondition inference (i.e., under which precondition, if violated, will the program always be incorrect?). We designed and implemented several new abstract interpretation-based analyses to infer necessary preconditions. The analyses infer atomic preconditions (including disjunctions), as well as universally and existentially quantified preconditions.

We experimentally validated the analyses on large scale industrial code.

For unannotated code, the inference algorithms find necessary preconditions for almost 64% of methods which contained warnings. In 27% of these cases the inferred preconditions were also sufficient, meaning all warnings within the method body disappeared. For annotated code, the inference algorithms find necessary preconditions for over 68% of methods with warnings. In almost 50% of these cases the preconditions were also sufficient. Overall, the precision improvement obtained by precondition inference (counted as the additional number of methods with no warnings) ranged between 9% and 21%.

6.11. Inference of Necessary Field Conditions with Abstract Interpretation

Participants: Mehdi Bouaziz, Manuel Fahndrich [Microsoft Research, Redmond, USA], Francesco Logozzo [Microsoft Research, Redmond, USA].

In [15], we present a new static analysis to infer necessary field conditions for object-oriented programs. A necessary field condition is a property that should hold on the fields of a given object, for otherwise there exists a calling context leading to a failure due to bad object state. Our analysis also infers the provenance of the necessary condition, so that if a necessary field condition is violated then an explanation containing the sequence of method calls leading to a failing assertion can be produced.

When the analysis is restricted to readonly fields, i.e., fields that can only be set in the initialization phase of an object, it infers object invariants. We provide empirical evidence on the usefulness of necessary field conditions by integrating the analysis into cccheck, our static analyzer for .NET.

Robust inference of readonly object field invariants was the #1 request from cccheck users.

6.12. TreeKs: A Functor to Make Numerical Abstract Domains Scalable

Participant: Mehdi Bouaziz.

Relational numerical abstract domains do not scale up. To ensure a linear cost of abstract domains, abstract interpretation-based tools analyzing large programs generally split the set of variables into independent smaller sets, sometimes sharing some non-relational information. In [14], we present a way to gain precision by keeping fully expressive relations between the subsets of variables, whilst retaining a linear complexity ensuring scalability.

6.13. An Abstract Domain to Infer Types over Zones in Spreadsheets

Participants: Cheng Tie, Xavier Rival.

abstract domains, spreadsheet script languages In [16], we proposed an abstract domain for the abstraction of spreadsheet contents.

Spreadsheet languages are very commonly used, by large user bases, yet they are error prone. However, many semantic issues and errors could be avoided by enforcing a stricter type discipline. As declaring and specifying type information would represent a prohibitive amount of work for users, we propose an abstract interpretation based static analysis for spreadsheet programs that infers type constraints over zones of spreadsheets, viewed as two-dimensional arrays. Our abstract domain consists in a cardinal power from a numerical abstraction describing zones in a spreadsheet to an abstraction of cell values, including type properties. We formalize this abstract domain and its operators (transfer functions, join, widening and reduction) as well as a static analysis for a simplified spreadsheet language. Last, we propose a representation for abstract values and present an implementation of our analysis.

6.14. Hierarchical Abstraction of Dynamic Structures

Participants: Pascal Sotin, Xavier Rival.

abstract domains, shape analysis, domain combination In [26], we designed a hierarchical shape abstract domain for the abstraction of complex data structures found in embedded softwares.

We propose a hierarchical shape abstract domain, so as to infer structural invariants of dynamic structures such as lists living inside static structures, such as arrays. This programming pattern is often used in safety critical embedded software that need to “allocate” dynamic structures inside static regions due to dynamic memory allocation being forbidden in this context. Our abstract domain precisely describes such hierarchies of structures. It combines several instances of simple shape abstract domains, dedicated to the representation of elementary shape properties, and also embeds a numerical abstract domain. This modular construction greatly simplifies the design and the implementation of the abstract domain. We provide an implementation, and show the effectiveness of our approach on a problem taken from a real code.

6.15. Reduced Product Combination of Abstract Domains for Shapes

Participants: Antoine Toubhans, Xavier Rival, Bor-Yuh Evan Chang [University of Colorado at Boulder].

abstract domains, shape analysis, reduced product In [27], we proposed a notion of reduced product for shape abstractions.

Real-world data structures are often enhanced with additional pointers capturing alternative paths through a basic inductive skeleton (e.g., back pointers, head pointers). From the static analysis point of view, we must obtain several interlocking shape invariants. At the same time, it is well understood in abstract interpretation design that supporting a separation of concerns is critically important to designing powerful static analyses. Such a separation of concerns is often obtained via a reduced product on a case-by-case basis. In this paper, we lift this idea to abstract domains for shape analyses, introducing a domain combination operator for memory abstractions. As an example, we present simultaneous separating shape graphs, a product construction that combines instances of separation logic-based shape domains. The key enabler for this construction is a static analysis on inductive data structure definitions to derive relations between the skeleton and the alternative paths. From the engineering standpoint, this construction allows each component to reason independently about different aspects of the data structure invariant and then separately exchange information via a reduction operator. From the usability standpoint, we enable describing a data structure invariant in terms of several inductive definitions that hold simultaneously.

ALF Project-Team

6. New Results

6.1. Processor Architecture within the ERC DAL project

Participants: Pierre Michaud, Nathanaël Prémillieu, Luis Germán Garcia Morales, Bharath Narasimha Swamy, Sylvain Collange, André Seznec, Arthur Pérais, Surya Narayanan, Sajith Kalathingal, Kamil Kedzierski.

Processor, cache, locality, memory hierarchy, branch prediction, multicore, power, temperature

Multicore processors have now become mainstream for both general-purpose and embedded computing. Instead of working on improving the architecture of the next generation multicore, with the DAL project, we deliberately anticipate the next few generations of multicores. While multicores featuring 1000s of cores might become feasible around 2020, there are strong indications that sequential programming style will continue to be dominant. Even future mainstream parallel applications will exhibit large sequential sections. Amdahl's law indicates that high performance on these sequential sections is needed to enable overall high performance on the whole application. On many (most) applications, the effective performance of future computer systems using a 1000-core processor chip will significantly depend on their performance on both sequential code sections and single threads.

We envision that, around 2020, the processor chips will feature a few complex cores and many (may be 1000's) simpler, more silicon and power effective cores.

In the DAL research project, http://www.irisa.fr/alf/dal, we explore the microarchitecture techniques that will be needed to enable high performance on such heterogeneous processor chips. Very high performance will be required on both sequential sections, —legacy sequential codes, sequential sections of parallel applications—, and critical threads on parallel applications, —e.g. the main thread controlling the application. Our research focuses essentially on enhancing single processes performance.

6.1.1. Microarchitecture exploration of control flow reconvergence

Participants: Nathanaël Prémillieu, André Seznec.

After continuous progress over the past 15 years [14], [13], the accuracy of branch predictors seems to be reaching a plateau. Other techniques to limit control dependency impact are needed. Control flow reconvergence is an interesting property of programs. After a multi-option control-flow instruction (i.e. either a conditional branch or an indirect jump including returns), all the possible paths merge at a given program point: the reconvergence point.

Superscalar processors rely on aggressive branch prediction, out-of-order execution and instruction level parallelism for achieving high performance. Therefore, on a superscalar core, the overall speculative execution after the mispredicted branch is cancelled, leading to a substantial waste of potential performance. However, deep pipelines and out-of-order execution induce that, when a branch misprediction is resolved, instructions following the reconvergence point have already been fetched, decoded and sometimes executed. While some of this executed work has to be cancelled since data dependencies exist, cancelling the control independent work is a waste of resources and performance. We have proposed a new hardware mechanism called SYRANT, SYmmetric Resource Allocation on Not-taken and Taken paths, addressing control flow reconvergence at a reasonable cost. Moreover, as a side contribution of this research we have shown that, for a modest hardware cost, the outcomes of the branches executed on the wrong paths can be used to guide branch prediction on the correct path [17].

As a follower work, we are now focusing on exploiting control flow reconvergence in the special case of predication. When the target ISA has predicated instruction, it is possible to transform control dependencies into data dependencies. This process is called if-conversion. As a result, the two paths of a conditional branch is merge into one path. Hence exploiting the principles developed in SYRANT is much easier than for a standard ISA.

6.1.2. Memory controller

Participant: André Seznec.

The memory controller has become one of the performance enablers of a computer system. Its impact is even higher on multicores than it was on uniprocessor systems. We propose the sErvice Value Aware memory scheduler (EVA) to enhance memory usage. EVA builds on two concepts, the request weight and the per-thread traffic light. For a read request on memory, the request weight is an evaluation of the work allowed by the request. Per-thread traffic lights are used to track whether or not in a given situation it is worth to service requests from a thread, e.g. if a given thread is blocked by refreshing on a rank then it is not worth to serve requests from the same thread on another rank. The EVA scheduler bases its scheduling decision on a service value which is heuristically computed using the request weight and per-thread traffic lights. Our EVA scheduler implementation relies on several hardware mechanisms, a request weight estimator, per-thread traffic estimators and a next row predictor. Using these components, our EVA scheduler estimates scores to issue scheduling decisions. EVA was shown to perform efficiently and fairly compared with previous proposed memory schedulers [21]

6.1.3. Performance and power models for heterogeneous muticores

Participants: Kamil Kedzierski, André Seznec.

In the DAL project, we expect architectures to be a combination of many simple cores for parallel execution and sequential accelerators [8] built on top of complex cores for ILP intensive tasks. For evaluating these architectures, we need performance and power models. We design a parallel manycore simulator, built with pthread implementation. Such an approach allows us to maintain flexibility and scalability: our goal is to scale well both when we vary the number of cores used to perform simulation, and as we vary the number of cores being simulated. Our implementation also allows to configure each core independently for the heterogeneous architectures. Preliminary results show that the simulator uses with very small memory footprint, which is crucial for the manycore studies with number of cores constantly increasing.

A new power management approach is needed for these future manycore processors that employ both sequential accelerators and simple cores. This is due to the fact that the frequency at which a given core operates is highly correlated with the cores' size (and thus a task that the core performs). Therefore, we built Dynamic Voltage Frequency Scaling model for the on-chip voltage regulator (VR) case, as we believe that future architectures will incorporate VRs on chip.

6.1.4. Designing supercores

Participants: Pierre Michaud, Luis Germán García Morales, André Seznec.

In the framework of the DAL project, we study super-cores that could achieve very high clock frequency and a high instruction per cycle rate (IPC). The current objective is to explore the design space of possible configurations for the microarchitecture that are suitable in terms of performance, area and power for the super-core. In particular, we focus on the back-end of the microarchitecture. A way to increase the IPC is to allow the core processing more instructions simultaneously e.g. increasing the issue width. This can be done for example by replicating the functional units (FU) inside the core. However keeping the same frequency could become very challenging. Clustering of FUs is a technique that helps designers to overcome this problem, even though other problems might appear e.g. IPC loss compared to an ideal monolithic back-end due to inter-cluster delays. We have started exploring different cluster schemes and instruction steering policies with the purpose of having a wide-issue clustered microarchitecture with a high IPC, a high frequency and the problem of inter-cluster delay minimized.

6.1.5. Helper threads

Participants: Bharath Narasimha Swamy, André Seznec.

Improving sequential performance will be key to both performance on single threaded codes and scalability on parallel codes. Complex out-of-order execution processors that aggressively exploit instruction level parallelism are the obvious design direction to improve sequential performance. However, ability of these complex cores to deliver performance will be undermined by performance degrading events such as branch mis-predictions and cache misses that limit the achievable instruction throughput. As an alternative to the monolithic complex core approach, we propose to improve sequential performance on emerging heterogeneous many core architectures by harnessing (unutilized) additional cores to work as helper cores for the sequential code. Helper cores can be employed to mitigate the impact of performance degrading events and boost sequential performance, for example by prefetching data for the sequential code ahead of time.

We are currently pursuing two directions to utilize helper cores. (1) We explore the use of helper cores to emulate prefetch algorithms in software. We will adapt and extend existing prefetch mechanisms for use on the helper cores and evaluate mechanisms to utilize both compute and cache resources on the helper cores to prefetch for the main thread. We intend to target delinquent load/store instructions that cause most of the cache misses and prefetch data ahead of time, possibly even before the hardware prefetchers on the main core. (2) We explore the use of helper cores to execute pre-computation code and generate prefetch requests for the main thread. Pre-computation code is constructed from the main thread and targets to capture the data access behavior of the main thread, particularly for irregular data access patterns in control-flow dominated code. We will explore algorithms to generate pre-computation code and evaluate mechanisms for communication and synchronization between the main thread and the helper cores, specifically in the context of a heterogenous many core architecture.

6.1.6. What makes parallel code sections and sequential code sections different?

Participants: Surya Natarajan, André Seznec.

In few years from now, single die processor components will feature many cores. They can be symmetric/asymmetric or homogeneous/heterogeneous cores. The utilization of these cores depends on the application and the programming model used. We have initiated a study on understanding the difference in nature between the parallel and sequential code sections in parallel applications. Initial experiments show that instruction mix of the serial and parallel parts are different. For example, contribution of the conditional branches are dominant in serial part and data transfer instructions are dominant in the parallel part. By experimentation, we infer that the conditional branch prediction in serial part needs a bigger branch predictor compared to the parallel part. Later, we would like to define the hardware mechanisms that are needed for cost effective execution of parallel sections; cost-effective meaning silicon and energy effective since parallelism can be leveraged.

On the other hand, the shared memory model has critical sections in the parallel sections, which makes the parallel sections sequential at times. We will try to characterize the nature of these sequential code sections and particularly identify their potential bottlenecks. The objective is to address the performance bottlenecks on sequential sections through new microarchitecture and/or compiler mechanisms.

6.1.7. Revisiting Value Prediction

Participants: Arthur Pérais, André Seznec.

Value prediction was proposed in the mid 90's to enhance the performance of high-end microprocessors. The research on Value Prediction techniques almost vanished in the early 2000's as it was more effective to increase the number of cores than to dedicate silicon to Value Prediction. However high end processor chips currently feature 8-16 high-end cores and the technology will allow to implement 50-100 of such cores on a single die in a foreseeable future. Amdahl's law suggests that the performance of most workloads will not scale to that level. Therefore, dedicating more silicon area to value prediction in high-end cores might be considered as worthwhile for future multicores.

We introduce a new value predictor VTAGE harnessing the global branch history [32]. VTAGE directly inherits the structure of the indirect jump predictor ITTAGE[11]. VTAGE is able to predict with a very high accuracy many values that were not correctly predicted by previously proposed predictors, such as the FCM predictor and the stride predictor. Three sources of information can be harnessed by these predictors: the global branch history, the differences of successive values and the local history of values. Moreover we show that the predictor components using these sources of information are all amenable to very high accuracy at the cost of some prediction coverage.

Compared with these previously proposed solutions, VTAGE can accommodate very long prediction latencies. The introduction of VTAGE opens the path to the design of new hybrid predictors. Using SPEC 2006 benchmarks, our study shows that with a large hybrid predictor, in average 55-60 % of the values can be predicted with more than 99.5 % accuracy. Evaluation of effective performance benefit is an on-going work.

6.1.8. Augmenting superscalar architecture for efficient many-thread parallel execution

Participants: Sylvain Collange, Sajith Kalathingal, André Seznec.

Heterogeneous multi-core architectures create many issues for test, design and optimizations. They also necessitate costly data transfer from the complex cores to the simple cores when switching from the parallel to sequential sections and vice-versa. We have initiated research on designing a unique core that efficiently run both sequential and massively parallel sections. It will explore how the architecture of a complex superscalar core has to be modified or enhanced to be able to support the parallel execution of many threads from the same application (10's or even 100's a la GPGPU on a single core). The overall objective is to support both sequential codes and very parallel execution, particularly data parallelism, on the same hardware core.

6.2. Other Architecture Studies

Participants: Damien Hardy, Pierre Michaud, Ricardo Andrés Velásquez, Sylvain Collange, André Seznec, Junjie Lai.

GPU, performance, simulation, vulnerability

6.2.1. Analytical model to estimate the performance vulnerability of caches and predictors to permanent faults

Participant: Damien Hardy.

This research was partially undertaken during Damien Hardy's stay in the Computer Architecture group of the University of Cyprus (January-August 2012).

Technology trends suggest that in tomorrow's computing systems, failures will become a commonplace due to many factors, and the expected probability of failure will increase with scaling. Faults can result in execution errors or simply in performance loss. Although faults can occur anywhere in the processor, the performance implications of a faulty cell vary depending on how the array is used in a processor.

Virtually all previous micro-architectural work aiming to assess the performance implications of permanently faulty cells relies on simulations with random fault-maps, assumes that faulty blocks are disabled, and focuses on architectural arrays such as caches.

These studies are, therefore, limited by the fault-maps they use that may not be representative for the average and distributed performance. Moreover, they are incomplete by ignoring faults in non-architectural arrays, such as predictors, that do not affect correctness but can degrade performance.

In [20], an analytical model is proposed for understanding the implications on performance of permanently faulty cells in caches and predictors. The model for a given program execution, micro-architectural configuration, and probability of cell failure, provides rapidly the *Performance Vulnerability Factor (PVF)*. PVF is a direct measure of the performance degradation due to permanent faults. In particular, the model can determine the expected PVF as well as the PVF probability distribution bounds without using an arbitrary number of random fault-maps.

The model, once derived, can be used to explore processor behavior with different cell probability of failures. This can be helpful to forecast how processor performance may be affected by faults in the future. Additionally, this information can be useful to determine which arrays have significant PVF and make design decisions to reduce their PVF, for example through a protection mechanism, using larger cells, or even by selecting a different array organization.

6.2.2. GPU-inspired throughput architectures

Participant: Sylvain Collange.

This research was partially undertaken while Sylvain Collange was with Universidade Federal de Minas Gerais, Belo Horizonte - Brazil, (January-September 2012).

In an heterogeneous architecture where power is the primary performance constraint, parallel sections of applications need to run on throughput-optimized cores that focus on energy efficiency. The Single-Instruction Multiple Thread (SIMT) execution model introduced for Graphics Processing Units (GPUs) provides inspiration to design such future energy-efficient throughput architectures. However, the performance of SIMT architectures is vulnerable to control and data flow divergences across threads. It limits its applicability to regular data-parallel applications. We work on making SIMT architectures more efficient, and generalizing the SIMT model to general-purpose architectures.

First, hybrids between multi-thread architectures and SIMT architectures can achieve a tradeoff between energy efficiency and flexibility [35]. Second, the same concepts that benefit GPUs may be applied to vectorize dynamically single-program, multi-thread applications. Indeed, data-parallel multi-thread workloads, such as OpenMP applications, expose parallelism by running many threads executing the same program. These threads may be synchronized to run the same instructions at the same time. SPMD threads also commonly perform the same computation on the same value. We take advantage from these correlations by sharing instructions between threads. It promises to save energy and frees processing resources on multi-threaded cores [26].

Besides architecture-level improvements, the efficiency of SIMT architectures can be improved through compiler-level code optimization. By maintaining a large number of threads in flight (in the order of tens of thousands), GPUs suffer from high cache contention as the local working set of each thread increases. This raises challenges as memory accesses are costly in terms of energy. Divergence analysis is a compiler pass that identifies similarities in the control flow and data flow of concurrent threads. In particular, it detects program variables that are affine functions of the thread identifier. Register allocation can benefit from divergence analysis to unify affine variables across SIMT threads and re-materialize them when needed. It reduces the volume of register spills, relieving pressure on the memory system [28].

6.2.3. Behavioral application-dependent superscalar core modeling

Participants: Ricardo Andrés Velásquez, Pierre Michaud, André Seznec.

Behavioral superscalar core modeling is a possible way to trade accuracy for processor simulation speed in situations where the focus of the study is not the core itself but what is outside the core, i.e., the *uncore*. In this modeling approach, a superscalar core is viewed as a black box emitting requests to the uncore at certain times. A behavioral core model can be connected to a cycle-accurate uncore model. Behavioral core models are built from detailed simulations. Once the time to build the model is amortized, significant simulation speedups are achieved.

We have proposed a new method for defining behavioral models for modern superscalar cores. Our method, *behavioral application-dependent superscalar core* (**BADCO**) modeling, requires two traces generated with cycle-accurate simulations to build a model. After the model is built, it can be used for simulating uncores. BADCO predicts the execution time of a thread running on a modern superscalar core with an error typically under 5%. From our experiments, we found that BADCO is qualitatively accurate, being able to predict how performance changes when we change the uncore. The simulation speedups obtained with BADCO are typically greater than 10 [29].

In a later work [33], we have shown that fast approximate microarchitecture models such as BADCO can also be very useful for selecting multiprogrammed workloads for evaluating the throughput of multicore processors. Computer architects usually study multiprogrammed workloads by considering a set of benchmarks and some combinations of these benchmarks. However, there is no standard method for selecting such sample, and different authors have used different methods. The choice of a particular sample impacts the conclusions of a study. Using BADCO, we propose and compare different sampling methods for defining multiprogrammed workloads for computer architecture [33]. We evaluate their effectiveness on a case study, the comparison of several multicore last-level cache replacement policies. We show that random sampling, the simplest method, is robust to define a representative sample of workloads, provided the sample is big enough. We propose a method for estimating the required sample size based on fast approximate simulation. We propose a new method, workload stratification, which is very effective at reducing the sample size in situations where random sampling would require large samples.

6.2.4. Performance Upperbound Analysis of GPU applications

Participants: Junjie Lai, André Seznec.

In the framework of the ANR Cosinus PetaQCD project, we are modeling the demands of high performance scientific applications on hardware. GPUs have become popular and cost-effective hardware platforms. In this context, we have been addressing the gap between theoretical peak performance on GPU and the effective performance [22]. There has been many studies on optimizing specific applications on GPU as well as and also a lot of studies on automatic tuning tools. However, the gap between the effective performance and the maximum theoretical performance is often huge. A tighter performance upperbound of an application is needed in order to evaluate whether further optimization is worth the effort. We designed a new approach to compute the CUDA application's performance upperbound through intrinsic algorithm information coupled with low-level hardware benchmarking. Our analysis [30] allows us to understand which parameters are critical to the performance and therefore to get more insight on the performance result. As an example, we analyzed the performance upperbound of SGEMM (Single-precision General Matrix Multiply) on Fermi and Kepler GPUs. Through this study, we uncover some undocumented features on Kepler GPU architecture. Based on our analysis, our implementations of SGEMM achieve the best performance on Fermi and Kepler GPUs so far (5 % improvement on average).

6.2.5. Multicore throughput metrics

Participant: Pierre Michaud.

Several different metrics have been proposed for quantifying the throughput of multicore processors. There is no clear consensus about which metric should be used. Some studies even use several throughput metrics. We have shown several new results concerning multicore throughput metrics [16]. We have exhibited the relation between single-thread average performance metrics and throughput metrics, emphasizing that throughput metrics inherit the meaning or lack of meaning of the corresponding single-thread metric [16]. In particular, two of the three most frequently used throughput metrics in microarchitecture studies, the weighted speedup and the harmonic mean of speedups, are inconsistent: they do not give equal importance to all benchmarks. We have demonstrated that the weighted speedup favors unfairness. We have shown that the harmonic mean of IPCs, a seldom used throughput metric, is actually consistent and has a physical meaning. We have explained under which conditions the arithmetic mean or the harmonic mean of IPCs can be used as strong indicators of throughput increase.

In a subsequent work [31], we have pointed out a problem with commonly used multiprogram throughput metrics, which is that they are based on the assumption that all the jobs execute for a fixed and equal time. We argue that this assumption is not realistic. We have proposed and characterized some new throughput metrics based on the assumption that jobs execute a fixed and equal quantity of work. We have shown that using such equal-work throughput metric may change the conclusion of a microarchitecture study [31].

6.3. Compiler, vectorization, interpretation

Participants: Erven Rohou, Emmanuel Riou, Arjun Suresh, André Seznec.

The usage of the bytecode-based languages such as Java has been generalized in the past few years. Applications are now very large and are deployed on many different platforms, since they are highly portable. With the new diversity of multicore platforms, functional, but also performance portability will become the major issue in the next 10 years. Therefore our research effort focuses on efficiently compiling towards bytecodes and on efficiently executing the bytecodes through JIT compilation or through direct interpretations.

6.3.1. Vectorization Technology To Improve Interpreter Performance

Participant: Erven Rohou.

Recent trends in consumer electronics have created a new category of portable, lightweight software applications. Typically, these applications have fast development cycles and short life spans. They run on a wide range of systems and are deployed in a target independent bytecode format over Internet and cellular networks. Their authors are untrusted third-party vendors, and they are executed in secure managed runtimes or virtual machines. Furthermore, due to security policies, these virtual machines are often lacking just-in-time compilers and are reliant on interpreter execution.

The main performance penalty in interpreters arises from instruction dispatch. Each bytecode requires a minimum number of machine instructions to be executed. In this work we introduce a powerful and portable representation that reduces instruction dispatch thanks to vectorization technology. It takes advantage of the vast research in vectorization and its presence in modern compilers. Thanks to a split compilation strategy, our approach exhibits almost no overhead. Complex compiler analyses are performed ahead of time. Their results are encoded on top of the bytecode language, becoming new SIMD IR (i.e., intermediate representation) instructions. The bytecode language remains unmodified, thus this representation is compatible with legacy interpreters.

This approach drastically reduces the number of instructions to interpret and improves execution time. SIMD IR instructions are mapped to hardware SIMD instructions when available, with a substantial improvement. Finally, we finely analyze the impact of our extension on the behavior of the caches and branch predictors.

These results are published in ACM TACO [18], and will be presented at the HiPEAC 2013 conference.

6.3.2. Tiptop

Participant: Erven Rohou.

Hardware performance monitoring counters have recently received a lot of attention. They have been used by diverse communities to understand and improve the quality of computing systems: for example, architects use them to extract application characteristics and propose new hardware mechanisms; compiler writers study how generated code behaves on particular hardware; software developers identify critical regions of their applications and evaluate design choices to select the best performing implementation.

We propose [27] that counters be used by all categories of users, in particular non-experts, and we advocate that a few simple metrics derived from these counters are relevant and useful. For example, a low IPC (number of executed instructions per cycle) indicates that the hardware is not performing at its best; a high cache miss ratio can suggest several causes, such as conflicts between processes in a multicore environment.

We propose tiptop: a new tool, similar to the UNIX top utility, that requires no special privilege and no modification of applications. Tiptop provides more informative estimates of the actual performance than existing UNIX utilities, and better ease of use than current tools based on performance monitoring counters. With several use cases, we have illustrated possible usages of such a tool.

Tiptop has been extended to display any user-defined arithmetic expression based on constants and counter values. A new configuration file lets users defined their default parameters as well as custom expressions.

6.3.3. Code obfuscation and JIT Compilers

Participant: Erven Rohou.

This project proposes to leverage JIT compilation to make software tamper-proof. The idea is to constantly generate different versions of an application, even while it runs, to make reverse engineering hopeless. A strong random number generator will guarantee that generated code is not reproducible, though the functionality is the same. Performance will not be sacrificed thanks to multi-core architectures: the JIT runs on separate cores, overlapping with the execution of the application.

The following directions are investigated:

1. We proposed a "change metric" that evaluates how different each new version of a function differs from the previous one, and hence contributes to the robustness of the system. The metric is based on string matching (such as in bioinformatics).
2. To increase the frequency of code switching, we consider on-stack-replacement. For performance, compilation is performed on a separate thread and pre-copying of the stack state to the new function version, thereby saving switching time.
3. We decompose a thread control-flow graph into many control-flow graphs such that the result of execution would be the same. The control-flow complexity is substantial as there are in the order of $O(n^n)$ possible combinations (where n is the number of threads and compilation units).

This is done in collaboration with the group of Prof. Ahmed El-Mahdy at E-JUST, Alexandria, Egypt.

6.3.4. Dynamic Analysis and Re-Optimization of Executables

Participants: Erven Rohou, Emmanuel Riou.

The objective of the ADT PADRONE beginning in November 2012 is to design and develop a platform for re-optimization of binary executables at run-time. We reviewed available support in hardware (such as performance monitoring unit, trap instructions), and in the Linux operating system (such as the ptrace system call). We started working on the platform, with an initial focus on analysis techniques.

6.3.5. Improving single core execution in the many-core era

Participants: Erven Rohou, André Seznec, Arjun Suresh.

In the framework of the DAL research project, we have initiated compiler research on using available unused resources in multicores to improve the performance of sequential code segments. Helper threads, driven by automated compiler infrastructure, can alleviate potential performance degradation due to resource contention. For example, loop based applications experiencing bad memory locality can be re-optimized by a just-in-time compiler to adjust to actual hardware characteristics.

6.4. WCET estimation

Participants: Damien Hardy, Benjamin Lesage, Hanbing Li, Isabelle Puaut, Erven Rohou, André Seznec.

Predicting the amount of resources required by embedded software is of prime importance for verifying that the system will fulfill its real-time and resource constraints. A particularly important point in hard real-time embedded systems is to predict the Worst-Case Execution Times (WCETs) of tasks, so that it can be proven that tasks temporal constraints (typically, deadlines) will be met. Our research concerns methods for obtaining automatically upper bounds of the execution times of applications on a given hardware. Our focus this year is on (i) multi-core architectures (ii) preemption delay analysis (iii) traceability of flow information in compilers for WCET estimation.

6.4.1. WCET estimation and multi-core systems

6.4.1.1. Predictable shared caches for mixed-criticality real-time systems

Participants: Benjamin Lesage, Isabelle Puaut, André Seznec.

The general adoption of multi-core architectures has raised new opportunities as well as new issues in all application domains. In the context of real-time applications, it has created one major opportunity and one major difficulty. On the one hand, the availability of multiple high performance cores has created the opportunity to mix on the same hardware platform the execution of a complex critical real-time workload and the execution of non-critical applications. On the other hand, for real-time tasks timing deadlines must be met and enforced. Hardware resource sharing inherent to multicores hinders the timing analysis of concurrent tasks. Two different objectives are then pursued: enforcing timing deadlines for real-time tasks and achieving highest possible performance for the non-critical workload.

In this work [23], we suggest a hybrid hardware-based cache partitioning scheme that aims at achieving these two objectives at the same time. Plainly considering inter-task conflicts on shared cache for real-time tasks yields very pessimistic timing estimates. We remove this pessimism by reserving private cache space for real-time tasks. Upon the creation of a real-time task, our scheme reserves a fixed number of cache lines per set for the task. Therefore uniprocessor worst case execution time (WCET) estimation techniques can be used, resulting in tight WCET estimates. Upon the termination of the real-time task, this private cache space is released and made available for all the executed threads including non-critical ones. That is, apart the private spaces reserved for the real-time tasks currently running, the cache space is shared by all tasks running on the processor, i.e. non-critical tasks but also the real-time tasks for their least recently used blocks. Experiments show that the proposed cache scheme allows to both guarantee the schedulability of a set of real-time tasks with tight timing constraints and enable high performance on the non-critical tasks.

6.4.1.2. WCET-oriented cache partitioning for multi-core systems

Participant: Isabelle Puaut.

Multi-core architectures are well suited to fulfill the increasing performance requirements of embedded real-time systems. However, such systems also require the capacity to estimate the timing behavior of their critical components. Interference between tasks, as they occur on standard multi-core micro-architectures due to cache sharing are still difficult to predict accurately. An alternative is to remove these indirect interferences between tasks through partitioning of the shared cache and through the use of partitioned task scheduling.

In this work [19], we have proposed a new algorithm for joint task and cache partitioning in multi-core systems scheduled using non-preemptive Earliest Deadline First policy. The main novelty of the algorithm is to take into account the tasks' period repartition in the task partitioning problem, which is critical in a non-preemptive context. Other task properties such as task cache requirements are also considered to optimize cache partitioning. Experiments show that our algorithm outperforms the state-of-the-art algorithm for tasks and cache partitioning, named IA3 [43], in terms of schedulability, specially when the spectrum of tasks periods is wide.

6.4.2. Preemption delay analysis for floating non-preemptive region scheduling

Participant: Isabelle Puaut.

This is joint work with Stefan M. Petters, Vincent Nélis and José Marinho, ISEP Porto, Portugal.

In real-time systems, there are two distinct trends for scheduling task sets on unicore systems: non-preemptive and preemptive scheduling. Non-preemptive scheduling is obviously not subject to any preemption delays but its schedulability may be quite poor, whereas fully preemptive scheduling is subject to preemption delays, but benefits from a higher flexibility in the scheduling decisions.

The time-delay involved by task preemptions is a major source of pessimism in the analysis of the task Worst-Case Execution Time (WCET) in real-time systems. Cache related preemption delays (CRPD) are the most important ones, and are caused by the preempting tasks that modify the cache; the preempted task then suffers an indirect delay after the preemption to reload the cache with useful information.

Preemptive scheduling policies including non-preemptive regions are a hybrid solution between non-preemptive and fully preemptive scheduling paradigms, which enables to conjugate both worlds benefits. In this work [25], we exploit the connection between the progression of a task in its operations, and the knowledge of the preemption delays as a function of its progression. Thus the pessimism in the preemption delay estimation is reduced, in comparison to state of the art methods, due to the increase in information available in the analysis. The method proposed in [25] was later improved in [24], to extract more information on the code and further decrease the CRPD estimation.

6.4.3. Traceability of flow information for WCET estimation

Participants: Hanbing Li, Isabelle Puaut, Erven Rohou.

Control-flow information is mandatory for WCET estimation, to guarantee that programs terminate (e.g. provision of bounds for the number of loop iterations) but also to obtain tight estimates (e.g. identification of infeasible or mutually exclusive paths). Such flow information is expressed though annotations, that may be calculated automatically by program/model analysis, or provided manually.

The objective of this work is to address the challenging issue of the mapping and transformation of the flow information from high level down to machine code. In a first step, we will consider only the issue of conveying information through the compilation flow, without any optimization. Then, we will study the impact of optimizations on the traceability of annotations.

This research started in October 2012 and is part of the ANR W-SEPT project.

AOSTE Project-Team

6. New Results

6.1. Logical time in Model-Driven Engineering embedded design

Participants: Charles André, Frédéric Mallet, Julien Deantoni, Marie-Agnès Peraldi Frati, Arda Goknil, Nicolas Chleq.

6.1.1. TimeSquare

We progressed our work on the foundations of logical time modeling as present in MARTE Time Model and our CCSL clock constraint specification language, while continuing the development of the TimeSquare tool environment which supports this in practice. A technical position paper was presented to the international TOOLS conference [22].

6.1.2. ECL (Event Constraint Language

Our contributions on CCSL and Time Model to the MARTE profile are part of the standard, but so far expressed in a syntax that is clearly distinct of the former UML notations. On the other hand, UML provides a textual language, named OCL, to express well-formedness constraints on diagram models and metamodels. While the original objectives were quite different, it seemed tempting to extend or adapt the general OCL philosophy, and to apply it then to timing and performance constraints as targeted by CCSL. The goal is to able the description of MoCs in an appropriate syntax, at metamodeling level. The result was a new syntax, called ECL for event contraint language, endowed with the well-established, sound timing interpretation as in CCSL. This work was reported in [40].

6.1.3. Logical time clocks to schedule data-flow models

Data-flow models can be used to capture data dependencies from applications, execution platforms and allocations. Most of the time such data dependencies impose only a partial order on the execution of application elements onto the execution platform and allow several allocation schemes. In [38], we have shown how to use logical time and CCSL constraints to capture explicitly the partial order imposed by the data-dependencies without imposing a total order. This work of representation expressivity then paved the way for analysis studies on time refinement, described in 6.3 .

6.1.4. Timing requirement modeling

One of the weak points of UML regarding a complete system design flow is its poor treatment of requirement capture (although this is partly corrected in the SysML profile). When requirements are made on timing aspects and logical time (as in our advocated approach), the relevant syntactic expressivity must be provided. We worked on the definition of a Domain-Specific Language (DSL for Timing Requirements engineering. The results were presented in [24], then applied to system specification in the context of the work described in section 6.6 .

6.2. Semantic translation of CCSL constraints into appropriate Büchi automata for trace recognition

Participants: Frédéric Mallet, Julien Deantoni, Robert de Simone, Ling Yin.

Our CCSL language expresses timing and scheduling constraints for a system, based on the notion of abstract logical clocks providing time events, and contraints linking them with relations of "asynchronous" nature (precedence, faster than) or of "synchronous" origin (subclocking, included in). Of course in a large system design both types coexist, and functional definitions also live next to declarative specifications to allow several timing solutions. Such a solution, called a schedule, must enforce that each logical clock either ticks endlessly, or terminates properly, in a way that globally respects the constraints. In previous works we have shown how a large variety of semantic scheduling constraints from the literature could soundly be represented in CCSL.

This year we focused on the semantic foundation of our CCSL language, by defining a structral operational semantic translation into a specific type of transition systems. Because we deal with infinite traces we had to consider acceptance mechanisms such as Büchi repeated states (as already used for translation of LTL temporal logic formulae in classical model-checking). Next we found out that, while state-labeled acceptance conditions were fine to obtain a direct and intuitive translation of individual constraints, building the composition of such models when dealing with multiple constraints was much easier in the case of *transition-labeled* Büchi automata (with repeated acceptance criteria now on transitions); the theory carries over to such case quite naturally, and has already been studied in the past. Finally, because traces must include infinite occurrences for *each* clock, we had to move to so-called *extended* Büchi automata, again a model already studied previously. We provided a complete semantic translation for all CCSL kernel constructs. Most importantly, we provided an efficient and simple fix-point algorithm to check the existance of a valid schedule, based on the type of automata just defined. This is (we believe) a genuine improvement on existing results, with potential applications outside our direct scope. These results are presented in a technical report, submitted for publication [46].

6.3. Timing refinement for multidimensional dataflow models using MARTE Time Model

Participants: Frédéric Mallet, Julien Deantoni, Jean-Vivien Millo.

Extensions of dataflow process networks have been proposed (as multidimensional SDF) to combine task parallelism (as in traditional process networks) with intensive data parallelism (as proposed in the Array-OL/Gaspard2 formalism developed in the DaRT EPI, for instance). The prospect of scheduling (seen as precise time cycle allocation) is here more complex, because of possible trade-offs between the granularity of treatments at task level *vs.* the size of data arrays that are handled uniformally in parallel inside each task. We considered how these phenomena could be represented (if not solved) inside the framework of MARTE Time Model and logical clocks, so as to handle such design issues in a well-defined MDE approach. Additionally, we used the MARTE platform description to specify how the previous models are refined through mapping allocation. The resulting modeling framework was presented in a journal article [19]. This work was conducted jointly with P. Boulet, from DaRT EPI, and C. Glitia, former DaRT PhD and Aoste postdoc student.

6.4. Process Network analysis

Participants: Robert de Simone, Jean-Vivien Millo.

6.4.1. K-periodic routing schemes for Network-on-Chip data traffic

This year we considered more specifically the issue of exploiting the predictable routing schemes of our KRG models, expressed as infinite binary words to indicate the sucessive branching directions at merge/select switch nodes, in order to encode data traffic patterns expanded at compile time, when mapping applications expressed under the form of dataflow process networks onto processor arrays in manycore architectures based on network-on-chip interconnects. To show the potential impact of such predicatble compile-time routing patterns, we stdudied as a typical example a fulll (all-to-all) broadcast algorithm on a mesh topology, connecting mode-less computation nodes as in the theory of cellular automata. This resulted in a precise recursive definition of routing patterns, which achieve an optimal data propagation (broadcast implemented as multicast), given the availability of actual links in the NoC topology. This result was presented at the Autamata'2012 conference [30], and an expanded version is available as technical report [44].

A wider view of the approach, and its potential benefits, are described in a technical report [43], submitted for publication.

6.4.2. Optimal data placement for process network scheduling

The topic of efficient scheduling of dataflow process network traffic to optimize both throughput and buffer queue sizing has given rise to a huge literature starting with seminal works in [49], [47], [56]. It has recently been given new impulse due to the advent of manycore architectures (see above). We conducted a number of theoretical works, to establish how such optimal computation scheduling can be best achieved in configurations where data are evenly distributed and streched in time across the (process) network. While this result is intuitively obvious, we formalized precisely what evenly distributed technically means, with the notion of balanced/mechanical words going a long way back in formal language theory, and we demonstrated that under such assumptions optimal schedules could be constructed *in a fully analytical way*, without any symbolic simulation steps or behavior expansion. The result was accepted for publication in a journal article [20].

6.5. Transformation from MARTE Time Model and CCSL to formal analysis models

Participants: Frédéric Mallet, Ling Yin.

This work was conducted in the context of an on-going collaboration with the Software Engineering Institute (SEI) of East Normal China University (ECNU) at Shanghai, which led altogether in part to the DAESD Associated-team, followed by a LIAMA joint project proposal recently submitted (HADES), and the co-supervision by Frédéric Mallet (together with Professor Jing Liu from ECNU) of the PhD thesis of Yin Ling. Yin Ling spent a one-year visit in our team, funded on a chinese governemental grant.

We studied the efficient and sound formal translation of a subset of CCSL contraints into the PROMELA/SPIN formalism, to benefit from model-checking formal analysis features in this environment. The translation is not completely direct, as synchronous simultaneity is not a native notion of PROMELA, and has to be encoded as atomicity. The motivating principles and translation details are provided in [42]. A similar attempt could be considered in the future, this time with the synchronous model-checker SMV, which allows coumpound instantaneous atomic behaviors.

Another line of research was initiated at ECNU to consider *logical continuous time*, while most of our current work considers only discrete time (while MARTE Time Model considers both). Considerations on *hybrid state diagrams* , inviting the expressive power of formal hierarchical hybrid automata models into the MDE design space of UML MARTE, were investigated in [27].

6.6. Use of MARTE Time Model and Logical Time in automotive design and AUTOSAR/TADL

Participants: Marie-Agnès Peraldi Frati, Julien Deantoni, Arda Goknil.

Precise timing constraint modeling and analysis [26], [33] is a key point for the correct development of automotive electronics. EAST-ADL and AUTOSAR has been adopted as standards in automotive industry. The timing model (TADL :Time augmented Description Language) of these standards raises different issues, mainly concerning the precise modeling of the multi clock characteristics of distributed systems together with parameterized timing expressions. In the ITEA TIMMO-2-USE project [35] 8.3.2.1 , we conducted a work [34], [35], on extending TADL with an explicit notion of multiple time bases for modeling the various temporal referentials used in an automotive design (clocks from different ECUs, motor position, etc). Additionally, timing constraints are augmented with parameters, which can be free at the highest abstraction level and then progressively defined during the design process. As a result, a symbolic timing expression in TADL2 is possibly made of a suitable set of arithmetic operators mixing symbolic identifiers (not necessarily set variables) and referring to different time bases. One typical use of this feature is to capture unknown configuration parameters for time budgeting; another one is to relate constraints in different time-bases to each other. Inherent to this work is also the study of the allowable ranges for symbolic values that are dictated by a set of constraints.

6.7. Multiview modeling and power intent in Systems-on-chip

Participants: Carlos Gomez Cardenas, Ameni Khecharem, Jean-François Le Tallec, Frédéric Mallet, Julien Deantoni, Robert de Simone.

6.7.1. High-level power management modeling

One of the concern of the UML MARTE profile is to allow non-functional property modeling, so that the same system bare description can be annotated in a number of views. In our case, combined with our logical time framework, such properties can be made as time-depending, inside potentially distinct views. We examplified this approach by dealing to a large extent with the example of low-power design and energy modeling in the case of Systems-on-Chip (SoC) in the mobile phone domain. Pure power/thermal modeling can be realized, based on the system global architecture, then made operational with the use of logical time controllers triggering power management functionalities.

Thermal/power simulation models are usually relying on continuous time. Therefore we considered the issue of *logical continuous* time, in an early attempt at combining simulation of continuous time power/thermal models with intrinsically discrete functional aspects. A prototype was realized in Scicos, as part of Ameni Khecharem master internship.

This work was conducted in the context of Carlos Gomez PhD thesis, and in collaboration with several partners inside the ANR HeLP project. It should be continued in the forthcoming PhD thesis of Ameni Khecharem, just started in the context of the follow-up ANR HOPE project, which will consider specific issues of hierachical power modeling and compositional power management (as an example of incremental multiview aspects).

6.7.2. IP-XACT

In this context of high-level power modeling and multiview concerns, we considered the emerging Accelera standard IP-XACT, made to provide easy-to-plug interfaces and Architecture Description Language (ADL) to allow simple assembly of hardware IP components into well-behaved SoCs. More specifically we provided means to annotate such interface with extra informations, directly borrowed from UML MARTE NFP properties, to handle power and thermal aspects. A number of model transformations back and forth between MARTE and (extended) IP-XACT were realized, and extraction of IP-XACT compliant interfaces from proprietary SystemC code describing the elementary IP component tehmselves has been defined and implemented as well.

This work was initiated as part of a project with STMicroelectronics, inside the nano2012 programme (ended 2011), and continued as part of the ANR HeLP collaboration. It resulted in the PhD thesis of Jean-François Le Tallec (who remained in the team for a couple of months later to complete the prototype implementation) [16].

6.8. Correct and efficient implementation of polychronous formalisms

Participants: Thomas Carle, Manel Djemal, Dumitru Potop Butucaru, Robert de Simone, Yves Sorel.

We extended our work on extending the AAA methodology for polychronous processes, by providing a better integration of clock analysis in the various phases of the implementation process (allocation, scheduling, pipelining, etc.). We also considered a wider range of implementation targets (time-triggered, MPSoC) and non-functional constraints (partitioning).

6.8.1. Time-Triggered Platform targets

Our first result this year concerns the automatic scheduling and code generation for time-triggered platforms. We extended our previous results in two significant ways. First, we designed a novel approach for specification of real-time features of time-triggered systems, with deadlines longer than periods; this allows a faithful representation of complex end-to-end flow requirements. Second, we provided new algorithms for off-line pipelined scheduling of these specifications onto partitioned time-triggered architectures *à la* ARINC 653; allocation of time slots/windows to partitions can be either complete or partially provided, or synthesized by our tool. Automatic allocation and scheduling onto multi-processor (distributed) systems with a global time base becomes feasible, taking into account communication costs. For single processors, we allow the generation of fully compliant ARINC653/APEX implementation code.

This work was mainly carried out inside the FUI Parsec 8.2.2.2 (which funds the PhD thesis of T. Carle) and P 8.2.2.1 projects, as well as a collaboration with ASTRIUM Space Transportation. First results are presented in a technical report, submitted for publication [39].

6.8.2. Multi-Processor System-on-Chip (MP-SoC) targets

Our second contribution concerns the automatic allocation and real-time scheduling over MPSoC (multi-processor on chip) architectures with NoC (network-on-chip) interconnect. One must take into account the specific 2D mesh network-on-chip topology, and synthesize the NoC routing patterns. This work provides operational execution support for the contributions described in 6.9 .

6.8.3. The LoPhT tool

Our recent work on extending the AAA methodology with better handling of execution conditions, with pipelining and pipelined scheduling, and with specific real-time scheduling and code generation techniques for time-triggered/partitioned and MPSoC platforms resulted in the development of a new scheduling and code generation toolbox, called LoPhT (for Logical to Physical Time Compiler).

6.9. Programmable On-Chip Networks

Participants: Thomas Carle, Manel Djemal, Dumitru Potop Butucaru, Robert de Simone, Zhen Zhang.

Modern computer architectures are increasingly relying on multi-processor systems-on-chip (MPSoCs), with data transfers between cores and memories managed by on-chip networks (NoC). This reflects in part a convergence between embedded, general-purpose PC, and high-performance computing (HPC) architecture designs.

Efficient compilation of applications onto MPSoCs remains largely an open problem, with the issue of best mapping of computation parts (threads, tasks,...) onto processing resources amply recognized, while the issue of best use of the interconnect NoC to route and transfer data still less commonly tackled. In the most general case, dynamic allocation of applications and channel virtualization can be guided by user-provided information under various forms, as in OpenMP, CUDA, OpenCL and so on. But then there is no clear guarantee of optimality, and first attempts by non-experts often show poor performances in the use of available computing power. Conversely there are consistent efforts, in the domains of embedded and HPC computing, aiming at automatic parallelization, compile-time mapping and scheduling optimization. They rely on the fact that applications are often known in advance, and deployed without disturbance from foreign applications, and without uncontrolled dynamic creation of tasks. Our contribution follows this “static application mapping” approach.

An optimal use of the NoC bandwidth should authorize data transfers to be realized according to (virtual) channels that are temporarily patterned to route data “just-in-time”. Previous works have identified the need for Quality of Service (QoS) in “some” data connections across the network (therefore borrowing notions from macroscopic networks, say internet and its protocols). But our experience with the AAA methodology strongly suggests that optimal NoC usage should result from a global optimization principle (embodied in a form of the AAA methodology), as opposed to a collection of local optimizations of individual connections. Indeed, various data flows with distinct sources and targets will nevertheless be highly concerted, both in time and space, like in a classical pipelined CPU, where the use of registers (replaced in our case with a complex NoC) is strongly synchronized with that of the functional units.

One main problem in applying such a global optimization approach is to provide the proper hardware infrastructures allowing the implementation of optimal computation and communication mappings and schedules. Our thesis is that optimal data transfer patterns should be encoded using simple programs configuring the router nodes (each router being then programmed to act its part in the global concerted computation and communication scheme).

We addressed this problem in the framework of our collaboration with the "Embedded Systems- on-Chips" department of the LIP6 laboratory, one of the main site of expertise for SoC/NoC design and Hardware/software codesign. This collaboration first materialized with the co-supervision of M. Djemal's PhD thesis. We concretely supported our proposed approach by extending the DSPIN 2D mesh network-on-chip (NoC) developed at UPMC- LIP6. In this NoC, we replace the fair arbitration modules of the NoC routers with static, micro-programmable modules that can enforce a given packet routing sequence, as specified by small programs. The design of such simple routing schemes can, for instance, be extracted from our results in section 6.4 .
We advocate the desired level of expressiveness/complexity for such simple configuration programs, and provide experimental data (cycle-accurate simulations) supporting our choices. We also wrote an architecture synthesis tool that allows simple architectural exploration of MPSoCs using the new DSPINPro NoC. First results in this direction have been presented in the DASIP 2012 conference, where our paper [23] has been short-listed for best paper award.

6.10. Uniprocessor Real-Time Scheduling

Participants: Laurent George, Mohamed Marouf, Daniel De Rauglaudre, Yves Sorel.

6.10.1. Combination of Non-Preemptive and Preemptive Tasks

We focused on fixed priority scheduling for a combination of non-preemptive strict periodic tasks in conjunction with preemptive sporadic tasks, that we extended to software fault tolerance [29]. We first investigated the transient phase for non-preemptive strict periodic tasks and we proved that its length is smaller than the transient phase for preemptive periodic tasks. Then, we determined the worst case scenario for preemptive sporadic tasks where the Worst Case Response Time (WCRT) can be obtained in the presence of strict periodic tasks. We proved that these release times belong only to the permanent phase of strict periodic tasks, and thus that the schedulability analysis for sporadic tasks can be restricted to the permanent phase. For preemptive sporadic tasks, we extended the classical necessary and sufficient schedulability condition based on the worst case response time computation to take into account non-preemptive strict periodic tasks. Finally, we considered software fault tolerance in the particular case where each primary strict periodic task has an alternate sporadic task which is released when the primary task fails. The schedulability analysis guarantees that even if all strict periodic tasks fail then all their respective alternate tasks will meet their deadlines.

6.10.2. Formal Proofs of Real-Time Scheduling Theorems

We completed two formal proofs of theorems in Coq on scheduling of fixed priority real-time preemptive tasks: one dealing with the sizes of busy periods (about 3500 lines of Coq), and another one dealing with response time (about 5200 lines of Coq). A monograph about these proofs, together with the formal check in Coq of scheduling conditions of strict periodicity, presented in the conference JFLA 2012 [37], have been started (currently about 70 pages).

6.11. Multiprocessor Real-Time Scheduling

Participants: Abderraouf Benyahia, Laurent George, Mohamed Marouf, Falou Ndoye, Simon Nivault, Yves Sorel, Cécile Stentzel, Meriem Zidouni.

6.11.1. Non-Preemptive Partitioned Fault Tolerant Scheduling

We addressed partitioned multiprocessor scheduling of non-preemptive strict periodic tasks which is extended thereafter to hardware fault tolerance [17].

In order to schedule a task set of non-preemptive strict periodic tasks on a multiprocessor platform, we partitioned this task set into subsets of tasks, each one is scheduled on a single processor using our proposed uniprocessor scheduling algorithm. The partition is carried out according to an enhanced “First Fit” algorithm that balances the load of the tasks on all the processors. However, inter-processors communications can lead to delay task execution. Thus, we determined the start time of each task taking into account the communication delay between this latter task and its predecessor tasks. Also, as inter-processor communications may generate a transient phase, we computed the length of the transient phase.

We proposed a fault tolerant real-time scheduling algorithm which allows hardware processors and/or buses faults, and conserves the strict periodicity of each task. We also proposed a graph transformation algorithm, applied on the task graph, which generates redundancies of tasks as well as dependencies. The transformation adds also selector tasks which choose data coming from the non failing processors and buses. That algorithm is based on exclusion relations to assign redundant tasks (resp. dependencies) to different processors (resp. busses). Then, we extended the previous partitioned multiprocessor scheduling algorithm to manage fault tolerance taking into account these exclusion relations.

This approach was successfully implemented on a CyCabs electric vehicle in a real-time fault tolerant tracking application where some processor or some bus could fail without any consequence on the proper execution of the application, i.e. same functional behaviour and real-time constraints satisfied.

6.11.2. Partitioned Scheduling with Exact Preemption Cost

Preemption allows a better scheduling success ratio but has a cost that must not be neglected in safety critical applications of domains such as avionic, automotive, etc. We focused on partitioned multiprocessor scheduling of independent preemptive periodic real-time tasks, while taking into account the exact preemption cost with the $\oplus$ operation formerly proposed by Meumeu and Sorel [10]. We improved the “greedy” heuristic proposed last year and compared it with the “Best-Fit” (BF) and “Worst-Fit” (WF) heuristics classically used in partitioned multiprocessor scheduling, but extended to take into account the exact preemption cost. We also compared our heuristic with an exact “Branch and Bound” algorithm with the same extension. The first comparison shows that the task allocation found by our heuristic gives a better response time than those found by WF and BF. This is due to the fact that the execution of the tasks is better parallelized. On the other hand, BF and WF heuristics execute a bit faster than our heuristic because they do not use all the available processors contrary to our heuristic which has the advantage to improve the load balancing of the tasks on all the processors.

Then, we addressed the scheduling of preemptive periodic real-time tasks with dependence constraints involving task precedences and data dependences. We considered harmonic tasks, i.e. periods of tasks are multiple or equal, to avoid loss of data. In order to satisfy data dependence constraints, we modified the release dates and deadlines of the dependent tasks according to the reception date of the data. In addition, data dependences between tasks mean to share data between dependent tasks which can cause deadlock and priority inversion problems. In order to solve these problems while taking into account the preemption cost, we proposed a new schedulability condition based on an extension of the $\oplus$ operation. We plan to propose a mutiprocessor scheduling heuristic based on that condition applied on tasks with modified release dates and deadlines.

6.11.3. Semi-partitioned Scheduling

Semi-partitioned multiprocessor scheduling stands between partitioned and global scheduling, the latter allowing migrations. We mainly addressed the semi-partitioned scheduling approach where the Worst Case Execution Time (WCET) of a job can be portioned, each portion being executed on a dedicated processor, according to a static pattern of migration. A job is migrated at its local deadline, computed from the deadline of the task it belongs to. We have studied this approach in the context of a fork/join task model with thread parallelism. A task is composed of a sequence of segments that can be parallelized in threads, if needed. The local deadlines depends on the number of parallel threads assigned to each segment.

6.11.4. Code Generation for Multicore

This work was carried out in the OPENPROD ITEA project 8.3.2.2 . xMod developed by IFPEN (IFP Energies Nouvelles), is an heterogeneous model integration environment that allows model importation from specific tools such as Simulink, AMSIM, etc. It also provides as a virtual instrumentation laboratory. In order to make xMod being able to run simulations with hardware-in-the-loop environment, we developed a new SynDEx executive kernel based on the kernel, dedicated to Windows/RTX, developed last year. That executive kernel is used with the macro-code generated by SynDEx to produce a real-time executable code that can drive the execution (real-time multi-core distribution and synchronized execution) of the models imported by xMod and simulated in the virtual instrumentation laboratory. This prototype as well as the report describing the corresponding achieved works, are the final deliverable of the OPENPROD project.

Furthermore, a French and English SynDEx code generation reference manual has been written to help future SynDEx users and maintainers to generate real-time code for already supported architectures or new ones.

6.11.5. Gateway with Modeling Languages for Certified Code Generation

This work was carried out in the P FUI project 8.2.2.1 . We provide inside the project expertise mainly on schedulability analysis and automatic generation of distributed real-time code. In this context, we developed a gateway between UML/MARTE and SynDEx. From a model specified with UML (Activity Diagram to specify algorithms and Composite Structure Diagram to specify multicomponent architectures) and refined with the UML profile MARTE (Modeling and Analysis of Real-Time Embedded Systems), we use the gateway to generate automatically distributed real-time application specified in the SynDEx format. Currently, we intend to provide a gateway between the GeneAuto language and SynDEx. The GeneAuto language is a subset of the future pivot P language. We presently deal with the part of the GeneAuto language corresponding to Simulink for data-flow modeling and we plan to deal soon with the part corresponding to Stateflow for control-flow modeling (composition of automata).

6.11.6. SynDEx Updates

We continued the software developments for the future version 8 of SynDEx which will feature a new software architecture to allow better functionality evolutions and maintenance. On the other hand in the COTROS ADT ("Génération de code temps réel distribué optimisé et sûr"), we completed the tests on the new automatic code generator for the current version 7 of SynDEx. This new generator produces code for mono-periodic and multi-periodic applications with condition and repetitive control structures, for the different hardware architectures supported by SynDEx. We developed a checker for the generated code that was integrated in the new generator. This checker verifies the correct use of semaphores and consequently the absence of deadlocks in the real-time code. Deadlocks are the most difficult part when dealing with distributed architectures. We achieved also a maintenance report describing the structure and the main features of code generator, as well as the technical choices we did.

6.12. Variability of program execution times on multicore processors

Participants: Sid-Ahmed-Ali Touati, Matias Vara Larsen, Abdelhafid Mazouz.

The activity described here represents the finalization of previous efforts conducted by Sid Touati and members of his groups, initiated before he joined the AOSTE EPI, and which are progressively merged with our own objectives, for results to be reported hopefully next year).

With the massive introduction of multicore platforms on embedded systems, parallel applications gained in performance. However, we showed in previous studies that the performance gain comes with high instability: program execution times vary in important way. We investigated the reasons for this variations and tried to understand the factors that influence program performance variability, that we decompose intro multiple families: factors from the application itself (implemented algorithms, coding technique, synchronization barriers, etc.), factors from the execution environment (OS effects, thread scheduling, Input/Ouput operations) and factors from the underlying hardware (micro-architecture, memory hierarchy, speculative execution, hardware data prefectching, etc.). Now, we have better understanding to these factors thanks to the work of two students:

1. Mr. Abdelhafid Mazouz who defended his PhD under the direction of Sid Touati at the university of Versailles in 11th of December 2012. The title of his PhD is "An Empirical Study of Program Performance of OpenMP Applications on Multicore Platforms".
2. Mr. Matias Vara Larsen, intern under the supervision of Sid Touati from February to June 2012, inside the Aoste EPI in Sophia-Antipolis, co-funded under a grant from Inria international internship program). The topic of his internship was to study the influence of he Linux kernels (multiple versions) on the stability of parallel applications.

Last, we published a rigorous statistical protocol in [21] called the Speedup-Test. It is used to analyze valid speedups (performance gain) in presence of performance instability: The Speedup-Test protocol is implemented and distributed as an open source tool based on R software. Our statistical methodology defines a consistent improvement compared with the usual performance analysis method in high-performance computing.

ARIC Team

6. New Results

6.1. Applications

Florent de Dinechin contributed high-performance signal processing on an FPGA to a prototype of high-throughput receiver for optical fiber transmission developed by Alcatel [33]. He also wrote a book chapter exposing the potential of FPGA-specific arithmetic for high-performance computing [49].

6.2. Hardware and FPGA Arithmetic

6.2.1. Mixed-precision fused multiply-and-add

With B. de Dinechin, from Kalray, N. Brunie and F. de Dinechin proposed to extend the classical fused-multiply-and-add operator with a larger addend and result. This enables higher-precision computation of sums of products at a cost that remains close to that of the classical FMA [29].

6.2.2. Multiplication by rational constants versus division by a constant

Motivated by the division by 3 or by 9 appearing in some stencil kernels, F. de Dinechin investigated how the periodicity of the binary representation of a rational constant could be exploited to design an architecture multiplying by this constant [18]. With L. S. Didier, this approach was then compared to a specialisation of divider architectures to the division by small integer constants, which is shown to match well the fine structure of FPGAs [32].

6.2.3. Floating-point exponentiation on FPGA

F. de Dinechin, with P. Echeverria and M. Lopez-Vallejo (U. Madrid) and B. Pasca (Altera), implemented the first floating-point unit for the pow and powr functions of the IEEE-754-2008 standard [50]. These functions compute x^y, and differ only in the specification of special cases. The implementation, parameterized in exponent and significand size, combines suitably modified exponential and logarithm units.

6.2.4. Arithmetic around the bit heap

F. de Dinechin, M. Istoan, G. Sergent, K. Illyes, B. Popa, and N. Brunie extended FloPoCo with a versatile framework for manipulating sums of weighted bits [51], [44]. This is a relevant way of implementing polynomials, filters and other coarse arithmetic cores.

6.2.5. Improving computing architectures

To improve High-Level Synthesis (HLS) for FPGAs, B. Pasca (former PhD student in AriC), with Ch. Alias (Inria Compsys) and A. Plesco (Zettice) developed tiling and scheduling algorithms that exploit the deeply pipelined operator at the core of a computing kernel [14].

With S. Collange and G. Diamos, N. Brunie proposed improvements in the architecture of general-purpose graphical processing units [28].

N. Brunie and F. de Dinechin, with Kalray's B. de Dinechin, are investigating embedding a reconfigurable core in the Kalray MPPA architecture. For this purpose, N. Brunie developed an environment for the design exploration of such an accelerator. This environment produces the hardware on one side, and its programming tools on the other side [43].

6.3. Elementary Functions

6.3.1. (M,p,k)-friendly points: a table-based method for trigonometric function evaluation

N. Brisebarre, M. Ercegovac (U. California at Los Angeles) and J.-M. Muller [25] present a new way of approximating the sine and cosine functions by a few table look-ups and additions. It consists in first reducing the input range to a very small interval by using rotations with "(M,p,k) friendly angles", proposed in this work, and then by using a bipartite table method in a small interval. An implementation of the method for 24-bit case is described and compared with CORDIC. Roughly, the proposed scheme offers a speedup of 2 compared with an unfolded double-rotation radix-2 CORDIC.

6.3.2. On Ziv's rounding test

With Ch. Lauter (LIP6), F. de Dinechin, J.-M. Muller and S. Torres proved and generalized a code sequence due to Ziv, which is used to round correctly a real value approximated (with a known error bound) as the unevaluated sum of two floating-point numbers [52].

6.4. Arithmetic Algorithms

6.4.1. Binary floating-point operators for VLIW integer processors

C.-P. Jeannerod and J. Jourdan-Lu [35] proposed software implementations of sinf, cosf and sincosf over [-pi/4, pi/4] that have proven 1-ulp accuracy and whose respective latencies on STMicroelectronics' ST231 VLIW integer processor are 19, 18 and 19 cycles. To get such performances they introduced a novel algorithm for simultaneous sine and cosine that combines univariate and bivariate polynomial evaluation schemes.

In the same context, C.-P. Jeannerod, J. Jourdan-Lu and C. Monat (STMicroelectronics Compilation Expertise Center, Grenoble) [36] studied the implementation of *custom* (i.e., specialized, fused, or simultaneous) operators, and provided qualitative evidence of the benefits of supporting such operators in addition to the five basic ones: this allows to be up to 4.2x faster on individual calls, and up to 1.59x faster on DSP kernels and benchmarks.

6.4.2. Error bounds for complex floating-point division with an FMA

Assuming that a fused multiply-add (FMA) instruction is available, C.-P. Jeannerod, N. Louvet and J.-M. Muller [37] obtained sharp error bounds for various alternatives to Kahan's 2 by 2 determinant algorithm. Combining such alternatives with Kahan's original scheme leads to componentwise-accurate algorithms for complex floating-point division, and for these algorithms sharp or reasonably sharp error bounds were also obtained.

6.4.3. Computation of correctly-rounded sums

P. Kornerup (U. of Southern Denmark), V. Lefèvre and J.-M. Muller [19] have shown that among the set of the algorithms with no comparisons performing only floating-point additions/subtractions, the 2Sum algorithm introduced by Knuth is minimal, both in terms of number of operations and depth of the dependency graph. They also prove that under reasonable conditions, an algorithm performing only round-to-nearest additions/subtractions cannot compute the round-to-nearest sum of at least three floating-point numbers. They also present new results about the computation of the correctly-rounded sum of three floating-point numbers.

6.4.4. Comparison between binary64 and decimal64 floating-point numbers

N. Brisebarre, C. Lauter (U. Paris 6), M. Mezzarobba and J.-M. Muller [27] introduce an algorithm that allows one to quickly compare a binary64 floating-point (FP) number and a decimal64 FP number, assuming the "binary encoding" of the decimal formats specified by the IEEE 754-2008 standard for FP arithmetic is used. It is a two-step algorithm: a first pass, based on the exponents only, makes it possible to quickly eliminate most cases, then when the first pass does not suffice, a more accurate second pass is required. They provide an implementation of several variants of their algorithm, and compare them.

6.5. Computer Algebra

6.5.1. Faster multivariate interpolation with multiplicities

M. Chowdhury (U. Western Ontario), C.-P. Jeannerod, V. Neiger (ENS de Lyon), É. Schost (U. Western Ontario) and G. Villard proposed fast randomized algorithms for interpolating multivariate polynomials with multiplicities. In the special bivariate case, this allows to accelerate the interpolation step of Guruswami and Sudan's list-decoding by a factor (list size)/(multiplicity).

6.5.2. On the complexity of solving quadratic boolean systems

M. Bardet (U. Rouen), J.-Ch. Faugère (PolSys), B. Salvy, and P.-J. Spaenlehauer (PolSys) [16] dealt with the fundamental problem in computer science of finding all the common zeroes of polynomials systems of quadratic polynomials over the field with 2 elements. The cryptanalysis of several modern ciphers reduces to this problem. Up to now, the best complexity bound was reached by an exhaustive search. They gave an algorithm that reduces the problem to a combination of exhaustive search and sparse linear algebra. This algorithm has several variants depending on the method used for the linear algebra step. Under precise algebraic assumptions, their complexity breaks the 2^n barrier. Experiments on random systems show that the algebraic assumptions are satisfied with probability very close to 1.

6.5.3. Power series solutions of singular (q)-differential equations

A. Bostan (Algorithms), M. F. I. Chowdhury (U. Western Ontario), R. Lebreton (Lix), B. Salvy, and É. Schost (U. Western Ontario) provided in [23] algorithms computing power series solutions of a large class of differential or q-differential equations or systems. Their number of arithmetic operations grows linearly with the precision, up to logarithmic terms.

6.5.4. Fast computation of common left multiples of linear ordinary differential operators

A. Bostan (Algorithms), F. Chyzak (Algorithms), Ziming Li (Chinese Academy of Sciences), and B. Salvy studied in [24] tight bounds and fast algorithms for LCLMs of several linear differential operators with polynomial coefficients. They analyzed the arithmetic complexity of existing algorithms for LCLMs, as well as the size of their outputs. They proposed a new algorithm that recasts the LCLM computation in a linear algebra problem on a polynomial matrix. This algorithm yields sharp bounds on the coefficient degrees of the LCLM, improving by one order of magnitude the best bounds obtained using previous algorithms. The complexity of the new algorithm is almost optimal, in the sense that it nearly matches the arithmetic size of the output.

6.5.5. Space complexity of fast D-finite function evaluation

M. Mezzarobba [41] showed that D-finite functions, i.e., solutions of linear differential equations with polynomial coefficients, can be evaluated in quasi-linear time and linear space with respect to the precision. In comparison, existing fast algorithms due to Chudnovsky and Chudnovsky and to van der Hoeven achieved the same time complexity with an overhead of a logarithmic factor in terms of memory usage.

6.5.6. Multiple precision evaluation of the Airy function with reduced cancellation

The series expansion at the origin of the Airy function $\mathrm{Ai}(x)$ is alternating and hence problematic to evaluate for $x > 0$ due to cancellation. Based on a method recently proposed by Gawronski, Müller, and Reinhard, Sylvain Chevillard and Marc Mezzarobba [31] exhibit two functions F and G, both with nonnegative Taylor expansions at the origin, such that $\mathrm{Ai}(x) = G(x)/F(x)$. The sums are now well-conditioned, but the Taylor coefficients of G turn out to obey an ill-conditioned three-term recurrence. They use the classical Miller algorithm to overcome this issue. They bound all errors and their implementation allows an arbitrary and certified accuracy, that can be used, e.g., for providing correct rounding in arbitrary precision.

6.5.7. Algorithms for combinatorial structures: well-founded systems and Newton iterations

C. Pivoteau (U. Marne-la-Vallée), B. Salvy, and M. Soria (UPMC) [21] considered systems of recursively defined combinatorial structures. They gave algorithms checking that these systems are well founded, computing generating series and providing numerical values. Their framework is an articulation of the constructible classes of Flajolet and Sedgewick with Joyal's species theory. They extend the implicit species theorem to structures of size zero. A quadratic iterative Newton method was shown to solve well-founded systems combinatorially. From there, truncations of the corresponding generating series were obtained in quasi-optimal complexity. This iteration transfers to a numerical scheme that converges unconditionally to the values of the generating series inside their disk of convergence. These results provide important subroutines in random generation. Finally, the approach was extended to combinatorial differential systems.

6.6. Euclidean Lattice Reduction and Applications

6.6.1. Lattice algorithms and hardness proofs

X.-W. Chang (McGill), D. Stehlé and G. Villard [17] proposed the first fully rigorous perturbation analysis of the R-factor of LLL-reduced matrices under column-wise perturbations. This study is very useful to devise LLL-type algorithms relying on floating-point approximations.

L. Luzzi (ENSEA), C. Ling (Imperial College) and D. Stehlé improved [20] the analyses of efficient Bounded Distance Decoding algorithms for lattices, and investigated the consequences for lattice-coded multiple-input multiple-output (MIMO) systems.

A. Langlois and D. Stehlé [54] introduced the Module-SIS and Module-LWE average-case lattice problems and reduced worst-case lattice problems to them. This provides a progressive transformation from the non-structured average-case lattices problems SIS and LWE, to the quite restricted but efficient average-case lattices problems Ring-SIS and Ring-LWE.

6.6.2. Cryptography

S. Ling (Nanyang Technological University, Singapore) and D. Stehlé [55] described the first public-key traitor tracing encryption scheme with security relying on the hardness of standard worst-case problems on Euclidean lattices.

J.-C. Belfiore (Telecom Paritech), L. Luzzi (ENSEA), C. Ling (Imperial College) and D. Stehlé [53] proved that nested lattice codes can achieve semantic security and strong secrecy over the Gaussian wiretap channel.

S. Ling (Nanyang Technological University, Singapore), K. Nguyen (NTU), H. Wang (NTU) and D. Stehlé [40] generalized Stern's zero-knowledge proof of knowledge protocol to obtain a statistical zero-knowledge proof of knowledge for the Inhomogeneous Small Integer Solution ISIS problem (in the infinity norm). This scheme is the first one that comes with no norm loss in the knowledge extraction procedure, leading to cryptographic constructions with tighter security proofs.

N. Attrapadung (AIST, Japan), J. Herranz (UPC, Spain), F. Laguillaumie, B. Libert (UCL, Belgium), E. de Panafieu (ENS Cachan), C. Ràfols (UPC, Spain) [15] proposed the first attribute-based encryption (ABE) schemes allowing for truly expressive access structures and with constant ciphertext size.

G. Castagnos (IMB) and F. Laguillaumie [38] gave a generic approach to design homomorphic encryption schemes, which extends Gjosteen's framework. A specific scheme allows an arbitrary number of multiplications in the groups, as well as a pairing evaluation on the underlying plaintexts.

J. Herranz (UPC, Spain), F. Laguillaumie, B. Libert (UCL, Belgium) and C. Ràfols (URV, Catalonia) [34] proposed the first two attribute-based (for threshold predicates) signature schemes with constant size signatures. Their security is proven in the selective-predicate and adaptive-message setting, in the standard model, under chosen message attacks.

S. Canard (Orange Labs), G. Fuchsbauer (University of Bristol, UK), A. Gouget (Gemalto), F. Laguillaumie [30] defined a new cryptographic primitive called plaintext-checkable encryption, which extends public-key encryption by the following functionality: given a plaintext, a ciphertext and a public key, it is universally possible to check whether the ciphertext encrypts the plaintext under the key. They provide efficient generic random-oracle constructions based on any probabilistic or deterministic encryption scheme as well as a practical construction in the standard model.

6.7. Reliability and Accuracy

6.7.1. Standardization of interval arithmetic

We contributed to the creation in 2008 and N. Revol chairs the IEEE 1788 working group on the standardization of interval arithmetic http://grouper.ieee.org/groups/1788/. More than 140 persons from over 20 countries take part in the discussions, around 1500 messages were exchanged in 2012. We are currently voting on portions of the text of the standard and have good hope that the group will reach a final version of the standard within the allotted time. An extension has been granted for 2 more years, until December 2014.

The annual in-person meeting, chaired by N. Revol, took place at the end of the SCAN 2012 conference in Novosibirsk, Russia, the 28th of September. It was broadcasted via the Web and feedback was possible through e-mails. More than 20 persons attended the meeting.

V. Lefèvre participated in various discussions, either in the mailing-list or in small subgroups (he sent around 390 mail messages in 2012). He proposed a motion, which passed, on properties needed by number formats for operations between intervals and numbers (constructors, midpoint, etc.).

The latest discussions dealt with:

- flavors: even if there continues to be a give-and-take between proponents of a “small” standard involving just basic interval arithmetic and those who also want to also include the less common “modal arithmetic”, this motion about “flavors” intends to allow inclusion of modal interval arithmetic consistently and simply, possibly at a later stage or revision of the standard;
- expressions: what is regarded as an expression by P1788, the relation with the programming languages, what this implies concerning the allowed optimizations, etc.;
- decorations: what are the properties of functions we want to track along a computation, how the empty interval is handled, etc.;
- reproducibility: across several runs of a translated (e.g., compiled) program or across platforms, representation-independent behavior, reproducibility for parallel programs, etc.

A personal view of the current status of the work of the IEEE P1788 group and of directions for future work has been presented in [46], [45].

6.7.2. Interval matrix multiplication

Several formulas exist for the product of two intervals using the midpoint-radius representation: they trade off accuracy for efficiency. The use of these formulas for the product of matrices with interval coefficients allows to use BLAS3 routines and to benefit from their performances in terms of execution time [48]. The accuracy of these methods are studied in [42]. As it can be difficult to ensure that a prescribed rounding mode is actually in use, formulas that are oblivious to the rounding mode are developed [22]. The implementations of these variants on multicores are compared in [47].

6.7.3. Rigorous polynomial approximation using Taylor models in Coq

One of the most common and practical ways of representing a real function on machines is by using a polynomial approximation. It is then important to properly handle the error introduced by such an approximation. N. Brisebarre, M. Joldes (Uppsala Univ., Sweden), E. Martin-Dorel, M. Mayero, J.-M. Muller, I. Pasca, L. Rideau (Marelle), and L. Théry (Marelle) have worked on the problem of offering guaranteed error bounds for a specific kind of rigorous polynomial approximation called Taylor model [26]. They carry out this work in the Coq proof assistant, with a special focus on genericity and efficiency for our implementation. They give an abstract interface for rigorous polynomial approximations, parameterized by the type of coefficients and the implementation of polynomials, and they instantiate this interface to the case of Taylor models with interval coefficients, while providing all the machinery for computing them.

ATEAMS Project-Team

5. New Results

5.1. Programming language support for statically type access to external resources

One of the open issues in programming is how to obtain typed access, including its beneficial IDE support, to data sources that have not been modeling with the programming language's data modeling facilities. Rather most data is modeled externally or not modeled at all. Mark Hills, Jurgen Vinju and Paul Klint proposed, designed and validated a programming language design where meta models for external data are imported and/or inferred at compile-time. These models are then used to generate source code to represent these models natively in the idiom of the programming language.

5.2. Statically analyzing PHP code

Tool support in IDEs for PHP code is limited due to the dynamic nature of the language. Mark Hills, Jurgen Vinju and Paul Klint produced a principled yet pragmatic infrastructure for analyzing PHP code nevertheless. The analyses first use crude but effective over-approximations of the PHP semantics to limit the search spaces and improve accuracy, then exploit information from user-manuals, and then use state-of-the-art static analysis techniques in a fixed point abstract interpretation to arrive at accurate results.

5.3. Modular Language Parametric Refactoring Framework

Anastasia Izmaylova with Jurgen Vinju produced a prototype implementation of a framework for specifying refactoring tools based on type constraints that is open to unpredicted language extensions. The problem with the co-evolution of programming language and their supporting refactoring tools is complexity. Often existing refactorings are not retro-fitted with the new language semantics and new opportunities for refactoring tools are not fulfilled. Anastasia designed a solution based on monad transformers that allows the kind of invasive extensibility needed to adapt complex existing implementation of language semantics with additional features that interact in many ways with the existing features.

5.4. Communication Action Emulation

CAE is a novel epistemic model for describing and evaluating the equivalence of communication models by Floor Sietsma and Jan van Eijck.

5.5. Notation-Parametric Grammar Recovery

Vadim Zaytsev generalized the algorithm for recovering context-free grammars from legacy language documentation. This facilitated the recovery of more grammars to be used in the study of grammarware and software language engineering.

5.6. (In)Validating Domain Knowledge Existence in Legacy Source Code

Davy Landman conducted a large experiment in comparing an extensive domain model to the information present in source code of applications that are used in the domain in question. The project management domain was chosen for this. The experiment is still in progress. Big steps were made in setting up the experiment, which includes reporting comprehensively on a large number of design decisions, in a traceable and reproducible manner.

5.7. Ensō

Ensōis a new programming system based on interpretation of domain-specific modeling languages. The system is co-designed and authors by Tijs van der Storm in collaboration with William Cook and Alex Loh. The two foundations of the Ensōsystem are managed data and object grammars. Managed data provides modular strategies for customizing how programming languages represent and provide access to data.

Object grammars form the second foundation: they facilate declarative, compositional, and bidirectional mappings from textual syntax to object graphs. Domain-specific models in Ensōare parsed and rendered using object grammars, and represented, in memory as managed data. Together they combine into a highly flexible and modular platform for model-driven development.

CAIRN Project-Team

6. New Results

6.1. Reconfigurable Architecture Design

6.1.1. Reconfiguration Controller

Participants: Robin Bonamy, Daniel Chillet, Sébastien Pillement.

Dynamically reconfigurable architectures, which can offer high performance, are increasingly used in different domains. Unfortunately, lots of applications cannot benefit from this new paradigm due to large timing overhead. Even for partial reconfiguration, modifying a small region of an FPGA takes few *ms* using the 14.5MB/s IP from Xilinx based on an embedded micro blaze processor. To cope with this problem by increasing performance, we have developed an ultra-fast power-aware reconfiguration controller (UPaRC) to boost the reconfiguration throughput up to 1.433 GB/s. UPaRC cannot only enhance the system performance, but also auto-adapt to various performance and consumption conditions. This could enlarge the range of supported applications and can optimize power-timing trade-off of reconfiguration phase for each selected application during run-time. The energy-efficiency of UPaRC over state-of-the-art reconfiguration controllers is up to 45 times more efficient [66].

6.1.2. Low-Power Reconfigurable Arithmetic Operators

Participants: Vivek D. Tovinakere, Olivier Sentieys, Arnaud Tisserand.

Arithmetic operators with fixed input data sizes are a source of unnecessary power consumption when data of lower precision have to be processed for significant amount of time. Configuring the arithmetic operator for lower precision when adequate and suppressing standby power in unused logic gates of the circuit can provide the benefit of reduced power consumption. In this work a logic clustering approach to partition arithmetic circuits as a function of reconfigurable input data widths is presented. Unused clusters at a specific precision are power-gated to achieve aggressive leakage power reduction that is a source of significant power consumption in nanoscale technologies. Application of this method to two types of 32-bit adders, reconfigurable to four precisions of data in 65nm CMOS technology shows a possible reduction in power consumption by a factor of 8 to 13 with an area overhead of 15% and 9.2% respectively. The variation of energy savings with respect to standby time of unused logic and frequency of precision adaptation was also analyzed.

6.1.3. Ultra-Low-Power Reconfigurable Controllers

Participants: Vivek D. Tovinakere, Olivier Sentieys, Steven Derrien.

Most digital systems use controllers based on a finite state machine (FSM) and datapath model. For specific control tasks, this model gives an energy efficient ASIC-like implementation compared to a microcontroller. This is especially true when the controller is required to execute a pre-specified task flow graph consisting of several basic tasks in applications like wireless sensor network (WSN) nodes. Previously design flows have been proposed to generate FSMs along with datapaths for tasks specified at a high level of abstraction and hence combine them with a scheduler to realize the overall controller. The generated controller was found to be efficient compared to its microcontroller counterpart by over two orders of magnitude in energy per operation metric, but a significant limitation of such controllers is the lack of flexibility. In this work, flexible controllers based on reconfigurable FSMs are considered at an expense of hardware area. Scalable architectures for reconfigurable FSMs based on lookup tables (LUTs) whose complexity may be parameterized by a high level specification of number of states, primary inputs and outputs of an FSM are proposed. Power gating as a low power technique is used to achieve aggressive leakage power reduction by shutting-off power to unused parts of logic at any given time. It is well known that in nanoscale CMOS circuits, the increase in static power density as a cost far exceeds the impact of area due to increased logic integration. The feedback and feedforward structures of a FSM are exploited to reduce programmable interconnections - a key issue in reconfigurable logic like FPGAs. Power estimation results show good performance of proposed architectures on different metrics when compared with other solutions in the design space of controllers for WSN nodes.

6.1.4. Models for Dynamically Reconfigurable Systems

6.1.4.1. Power Models

Participants: Robin Bonamy, Daniel Chillet, Olivier Sentieys.

Including a reconfigurable area in a heterogeneous system-on-chip is considered as an interesting solution to reduce area and increase performance. But the key challenge in the context of embedded systems is currently the power budget of the system, and the designer needs some early estimations of the power consumption of its system. Power estimation for reconfigurable systems is a difficult problem because several parameters need to be taken into account to define an accurate model.

In this work, we considered dynamic reconfiguration that makes possible to partially reconfigure a specific part of the circuit while the rest of the system is running. This technique has two main effects on power consumption. First, thanks to the area sharing ability, the global size of the device can be reduced and the static (leakage) power consumption can thus be also reduced. Secondly, it is possible to delete the configuration of a part of the device which reduces the dynamic power consumption when a task is no longer used. We have defined several models of power consumption for the dynamic reconfiguration on a Virtex 5 board and a first model of the power consumption of the reconfiguration. This model shows that the power consumption not only depends on the bitstream file size but also on the content of the reconfiguration region. Finally three models of the partial and dynamic reconfiguration with different complexities/accuracy tradeoffs are extracted [52].

6.1.4.2. High-Level Modeling of Reconfigurable Architectures

Participants: Robin Bonamy, Daniel Chillet.

To model complex multiprocessor SoCs, the Architecture Analysis & Design Language (AADL) has been adopted. We have proposed an extension of AADL towards reconfigurable systems to support power consumption and dynamic reconfiguration modeling. As different power/energy/time/cost tradeoffs can be achieved for a given application, we proposed to represent as Pareto frontiers the set of values of power/energy vs. execution time or cost to model the execution of an application on the reconfigurable system. These Pareto frontiers are computed from analysis functions which extract and combine component characteristics from AADL models. These functions, developed in OCL (Object Constraint Language), are well suited for design space exploration and they can be used to extract the energy/power properties from the model to compute and to verify user's constraints.

To complete these levels of description, we started the development of techniques for constraint verifications. These developments are based on the OCL language, which allows one to extract characteristics on the AADL model, compute mathematical expressions and finally verify mathematical constraints. These verifications have been developed for power and energy consumption, they include static and dynamic power estimation, the power consumption during the dynamic reconfiguration process and the reconfiguration speed. They handle all energy/power parameters related to reconfigurable architectures for an energy estimation of a complete application and heterogeneous system. We currently work on the link between the design space exploration explained in the previous section and the AADL models developed in collaboration with the LEAT laboratory, and to be included in the Open-People Platform [27], [54], [76], [71].

6.1.5. Fault-Tolerant Reconfigurable Architectures

Participants: Sébastien Pillement, Manh Pham, Stanislaw Piestrak [Univ. Metz].

In terms of complex systems implementation, reconfigurable FPGAs circuits are now part of the mainstream thanks to their flexibility, performance and high number of integrated resources. FPGAs enter new fields of applications such as aeronautics, military, automotive or confined control thanks to their ability to be remotely updated. However, these fields of applications correspond to harsh environments (cosmic radiation, ionizing, electromagnetic noise) and with high fault-tolerance requirements. We proposed a complete framework to design reconfigurable architecture supporting fault-tolerance mitigation schemes. The proposed framework enables simulation, validation of mitigation operations, but also the scaling of architecture resources. The proposed model was validated thanks to a physical implementation of the fault-tolerant reconfigurable platform. Results have shown the effectiveness of the framework [39] and confirmed the potential of dynamically reconfigurable architectures for supporting fault-tolerance in embedded systems.

6.1.6. Low-Power Architectures

6.1.6.1. Wakeup Time and Wakeup Energy Estimation in Power-Gated Logic Clusters

Participants: Olivier Sentieys, Vivek D. Tovinakere.

Run-time power gating for aggressive leakage reduction has brought into focus the cost of mode transition overheads due to frequent switching between sleep and active modes of circuit operation. In order to design circuits for effective power gating, logic circuits must be characterized for overheads they present during mode transitions. We have proposed a method to determine steady-state virtual-supply voltage in active mode and hence present a model for virtual-supply voltage in terms of basic circuit parameters. Further, we derived expressions for the estimation of two mode transition overheads: wakeup time and wakeup energy for a power-gated logic cluster using the proposed model. Experimental results of application of the model to ISCAS85 benchmark circuits show that wakeup time may be estimated within a low average error across large variation in sleep transistor sizes and variation in circuit sizes with significant speedup in computation time compared to transistor-level circuit simulations [73].

6.1.7. Arithmetic Operators for Cryptography

Participants: Arnaud Tisserand, Emmanuel Casseau, Thomas Chabrier, Danuta Pamula, Karim Bigou, Franck Bucheron, Jérémie Métairie.

6.1.7.1. Arithmetic Operators for Fast and Secure Cryptography

Electrical activity variations in a circuit are one of the information leakage used in side channel attacks. In [65], we present $\mathbb{F}_{2^m}$ finite-field multipliers with reduced activity variations for asymmetric cryptography. Useful activity of typical multiplication algorithms is evaluated. The results show strong shapes, which can be used as a small source of information leakage. We propose modified multiplication algorithms and architectures to reduce useful activity variations. Useful activity has been evaluated using accurate FPGA emulation and activity counters at every operation cycle. Measurement analysis shows that the implemented multiplication algorithms (classical, Montgomery and Mastrovito) lead to specific shapes for the curve of activity variations which may be used as a small source of information leakage for some side channel attacks. We proposed modifications of selected $\mathbb{F}_{2^m}$ multipliers to reduce this information leakage source at two levels: architecture level by removing activity peaks due to control (e.g. reset at initialization) and algorithmic level by modifying the shape of the activity variations curve. Due to very low-level optimizations there is no significant area and delay overhead.

Paper [64] presents overview of the most interesting $\mathbb{F}_{2^m}$ multiplication algorithms and proposes efficient hardware solutions applicable to elliptic curve cryptosystems. It focuses on fields of size $m = 233$, one of the sizes recommended by NIST (National Institute of Standards and Technology). We perform an analysis of most popular algorithms used for multiplication over finite fields; suggest efficient hardware solutions and point advantages and disadvantages of each algorithm. The article overviews and compares classic, Mastrovito and Montgomery multipliers. Hardware solutions presented here, implement their modified versions to gain on efficiency of the solutions. Moreover we try to present a fair comparison with existing solutions. The designs presented here are targeted to FPGA devices.

6.1.7.2. ECC Processor with Protections Against SCA

A dedicated processor for elliptic curve cryptography (ECC) is under development. Functional units for arithmetic operations in $\mathbb{F}_{2^m}$ and $\mathbb{F}_p$ finite fields and 160–600-bit operands have been developed for FPGA implementation. Several protection methods against side channel attacks (SCA) have been studied. The use of some number systems, especially very redundant ones, allows one to change the way some computations are performed and then their effects on side channel traces.

6.1.8. 3D Heterogeneous SoC Design

Participants: Quang-Hai Khuat, Hoa Le, Sébastien Pillement, Emmanuel Casseau, Antoine Courtay, Daniel Chillet, Olivier Sentieys.

A three-dimensional system-on-chip is an SoC in which two or more layers of dies are stacked vertically into a single circuit and integrated within a single package. 3D stacking is an emerging solution that provides a new dimension in performance by reducing the distances that signals need to travel between the different blocks of a system. Interconnects in future technologies are known to be a major bottleneck for performance and power. In this context, 3D implementations can help alleviate the performance and power overheads of on-chip wiring.

In the context of 3D SoC, we have developed a spatio-temporal scheduling algorithm for 3D architecture composed of two layers: i) a homogenous Chip MultiProcessor (CMP) layer and ii) a homogeneous embedded Field-Programmable Gate Array (eFPGA) layer, interconnected by through-silicon vias (TSVs), thus ensuring tight coupling between software tasks on processors and associated hardware accelerators on the eFPGA. We extended the Proportionate-fair (Pfair) algorithm to tackle 3D heterogeneous multiprocessors. Unlike Pfair, our algorithm copes with task dependencies and global communication cost. Communication cost is computed by summing not only point-to-point/direct communication cost, but also memory cost. Our algorithm favours direct communication onto the eFPGA layer, but uses shared memory when direct communications are not possible [61], [75], [74].

6.2. Compilation and Synthesis for Reconfigurable Platform

Participants: Steven Derrien, Emmanuel Casseau, Daniel Menard, François Charot, Christophe Wolinski, Olivier Sentieys, Patrice Quinton.

6.2.1. Polyhedral-Based Loop Transformations for High-Level Synthesis

Participants: Steven Derrien, Antoine Morvan, Patrice Quinton.

After almost two decades of research effort, there now exists a large choice of robust and mature C to hardware tools that are used as production tools by world-class chip vendor companies. Although these tools dramatically slash design time, their ability to generate efficient accelerators is still limited, and they rely on the designer to expose parallelism and to use appropriate data layout in the source program. We believe this can be overcome by tackling the problem directly at the source level, using source-to-source optimizing compilers. More precisely, our aim is to study how polyhedral-based program analysis and transformation can be used to address this problem. In the context of the PhD of Antoine Morvan, we have studied how it was possible to improve the efficiency and applicability of nested loop pipelining (also known as nested software pipelining) in C to hardware tools. Loop pipelining is a key transformation in high-level synthesis tools as it helps maximizing both computational throughput and hardware utilization. Nevertheless, it somewhat looses its efficiency when dealing with small trip-count inner loops, as the pipeline latency overhead quickly limits its efficiency. Even if it is possible to overcome this limitation by pipelining the execution of a whole loop nest, the applicability of nested loop pipelining has so far been limited to a very narrow subset of loops, namely perfectly nested loops with constant bounds. In this work, we have extended the applicability of nested-loop pipelining to imperfectly nested loops with affine dependencies. We have shown how such loop nest can be analyzed and, under certain conditions, how one can modify the source code in order to allow nested loop pipeline to be applied using a method called polyhedral bubble insertion. The approach has been implemented in the Gecos source-to-source toolbox and was validated using two leading-edge HLS commercial tools. It helps improving performance for a minor area overhead. This work has been accepted for publication in late 2012 to IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. In addition, the complete Gecos source-to-source toolbox was presented at the DAC university booth in June 2012.

In addition to our work on nested loop pipelining, we also started investigating how to extend existing polyhedral code generation technique to enable the synthesis of area-efficient control-logic for nested loops hardware accelerators.

6.2.2. Compiling for Embedded Reconfigurable Multi-Core Architectures

Participants: Steven Derrien, Olivier Sentieys, Maxime Naullet.

Current and future wireless communication and video standards have huge processing power requirements, which cannot be satisfied with current embedded single processor platforms. Most platforms now therefore integrate several processing core within a single chip, leading to what is known as embedded multi-core platforms. This trend will continue, and embedded system design will soon have to implement their systems on platforms comprising tens if not hundred of high performance processing cores. Examples of such architectures are the Xentium processor from by Recore or the Kahrisma processor, a radically new concept of morphable processor from Karlsruhe Institute of Technology (KIT). This evolution will pose significant design challenges, as parallel programming is notoriously difficult, even for domain experts. In the context of the FP7 European Project Alma (Architecture-oriented parallelization for high performance embedded Multi-core systems using scilAb), we are studying how to help designers programming these platforms by allowing them to start from a specification in Matlab and/or Scilab, which are widely used for prototyping image/video and wireless communication applications. Our research work in this field revolves around two topics. The first one aims at exploring how floating-point to fixed-point conversion can be performed jointly with the SIMD instruction selection stage to explore performance/accuracy trade-off in the software final implementation. The second one aims at exploring how program transformation techniques (leveraging the polyhedral model and/or based on the domain specific semantics of scilab built-in functions) can be used to enable an efficient coarse grain parallelization of the target application on such multi-core machines.

6.2.3. Reconfigurable Processor Extensions Generation

Participants: Christophe Wolinski, François Charot, Antoine Floc'h.

Most proposed techniques for automatic instruction sets extension usually dissociate pattern selection and instruction scheduling steps. The effects of the selection on the scheduling subsequently produced by the compiler must be predicted. This approach is suitable for specialized instructions having a one-cycle duration because the prediction will be correct in this case. However, for multi-cycle instructions, a selection that does not take into account scheduling is likely to privilege instructions which will be, a posteriori, less interesting than others in particular in the case where they can be executed in parallel with the processor core.

The originality of our research work is to carry out specialized instructions selection and scheduling in a single optimization step. This complex problem is modeled and solved using constraint programming. This approach allows the features of the extensible processor to be taken into account with a high degree of flexibility. Two architecture models are envisioned. The first one is an extensible processor tightly coupled to an hardware extension having internal registers used to store intermediate results. The second model is VLIW-oriented, a specialized instruction is able to configure several processing using working in parallel. Our experimental results show that these approaches are able to handle graphs of several hundred of nodes in a reasonable time (less than ten seconds for most cases). Speedups obtained are particularly interesting for applications having a high degree of instruction-level parallelism.

More details on constraint programming approach applied to reconfigurable processor extension generation can be found in [32] and in the Ph.D. thesis of Antoine Floc'h [20].

During this year, we have also studied a novel technique that addresses the interactions between code optimization and instruction set extension. The idea is to automatically transform the original loop nests of a program (using the polyhedral model) to select specialized and vectorizable instructions. These instructions may use local memories of the hardware extension to store intermediates data produced at a given loop iteration. Details can be found in the Ph.D. thesis of Antoine Floc'h [20].

6.2.4. Custom Operator Identification for High-Level Synthesis

Participants: Emmanuel Casseau, François Charot, Chenglong Xiao.

In this work, our goal is to propose an automated design flow based on custom operator identification for high-level synthesis. Custom operators that can be implemented in special hardware units make it possible to improve performance and reduce area of the design. The key issues involved in the design flow are: automatic enumeration and selection of custom operators from a given high-level application code and re-generation of the source code incorporating the selected custom operators. This new source code is then provided to the high-level synthesis tool. The application is first translated into an internal representation based on a graph representation. Then the problem is to enumerate and select subgraphs that will be implemented as custom operators. However, enumerating all the subgraphs is a computationally difficult problem. In Xiao's PhD thesis [25] and [42], three enumeration algorithms for exact enumeration of subgraphs under various constraints were proposed. Compared to a previously proposed well-known algorithm, the proposed enumeration algorithms can achieve orders of magnitude speedup. Selecting a most profitable subset from the enumerated subgraphs is also a time-consuming job. [25] proposed three different selection heuristics targeting different objectives. Based on these algorithms, experimental results show that the approach achieves on average 19% area reduction, compared to a traditional high-level synthesis with CtoS tool from Cadence. Meanwhile, the latency is reduced on average by 22%.

6.3. Interaction between Algorithms and Architectures

6.3.1. Numerical Accuracy Analysis and Optimization

Participants: Daniel Menard, Karthick Parashar, Olivier Sentieys, Romuald Rocher, Pascal Scalart, Aymen Chakhari, Jean-Charles Naud, Emmanuel Casseau.

Most of analytical methods for numerical accuracy evaluation use perturbation theory to provide the expression of the quantization noise at the output of a system. Existing analytical methods do not consider a correlation between noise sources. This assumption is no longer valid when a unique datum is quantized several times. In [34], an analytical model of the correlation between quantization noises is provided. The different quantization modes are supported and the number of eliminated bits is taken into account. The expression of the power of the output quantization noise is provided when the correlation between the noise sources is considered. The proposed approach allows improving significantly the estimation of the output quantization noise power compared to the classical approach, with a slight increase of the computation time.

An analytical approach is studied to determine accuracy of systems including unsmooth operators. An unsmooth operator represents a function which is not derivable in all its definition interval (for example the sign operator). The classical model is no valid yet since these operators introduce errors that do not respect the Widrow assumption (their values are often higher than signal power). So an approach based on the distribution of the signal and the noise is proposed. It is applied to the sphere decoding algorithm to determine analytically the error probability due to quantization [53]. We also focus on recursive structures where an error influences future decision. So, the Decision Feedback Equalizer is also considered. In that case, numerical analysis method (as Newton Raphson algorithm) can be used. Moreover, an upper bound of the error probability can be analytically determined. A method to determine the distribution of the noise due to quantization at the output of a system made of smooth operators has been developed [70]. It is based on Generalized Gaussian Distribution and allows take under consideration all possible distributions (uniform, gaussian, laplacian, etc.).

6.3.2. Multi-Antenna Systems

Participants: Olivier Berder, Pascal Scalart, Quoc-Tuong Ngo, Viet-Hoa Nguyen.

Still considering the maximization of the minimum Euclidean distance, we proposed a new linear precoder obtained by observing the SNR-like precoding matrix. An approximation of the minimum distance is derived, and its maximum value was obtained by maximizing the minimum diagonal element of the SNR-like matrix. The precoding matrix is first parameterized as the product of a diagonal power allocation matrix and an input-shaping matrix acting on rotation and scaling of the input symbols on each virtual subchannel. We demonstrated that the minimum diagonal entry of the SNR-like matrix is obtained when the input-shaping matrix is a DFT-matrix. The major advantage of this design is that the solution can be available for all rectangular QAM-modulations and for any number of datastreams [35], [36], [37]. To reduce the decoding complexity of linearly precoded MIMO systems, the sphere decoder was applied instead of maximum likelihood and the performance complexity trade-off was investigated. The sphere decoding (SD) algorithm, proposed as a sub-optimal ML-decoding, just considers a subset of lattice points that drop into the sphere centered by the received point to obtain the decoded solution, thus reducing significantly the complexity. Because the structure of our precoder is complicated and strongly depends on the channel, it exists the case when all power is poured only on the best sub-channel. Some adjustments, therefore, of traditional sphere decoding algorithm were mandatory to adapt to the precoded MIMO systems.

6.3.3. Impact of RF Front-End Nonlinearity on WSN Communications

Participants: Amine Didioui, Olivier Sentieys, Carolynn Bernier [CEA Leti].

6.3.4. HarvWSNet: A Co-Simulation Framework for Energy Harvesting Wireless Sensor Networks

Participants: Amine Didioui, Olivier Sentieys, Carolynn Bernier [CEA Leti].

Recent advances in energy harvesting (EH) technologies now allow wireless sensor networks (WSNs) to extend their lifetime by scavenging the energy available in their environment. While simulation is the most widely used method to design and evaluate network protocols for WSNs is simulation, existing network simulators are not adapted to the simulation of EH-WSNs and most of them provide only a simple linear battery model. To overcome these issues, we have proposed HarvWSNet, a co-simulation framework based on WSNet and Matlab that provides adequate tools for evaluating EH-WSN lifetime [56]. Indeed, the framework allows for the simulation of multi-node network scenarios while including a detailed description of each node's energy harvesting and management subsystem and its time-varying environmental parameters. A case study based on a temperature monitoring application has demonstrated HarvWSNet?s ability to predict network lifetime while minimally penalizing simulation time.

6.3.5. Cooperative Strategies for Low-Energy Wireless Networks

Participants: Olivier Berder, Olivier Sentieys, Le-Quang-Vinh Tran, Duc-Long Nguyen.

Recently, cooperative relay techniques (e.g. repetition-based or distributed space-time code based (DSTC-based) protocols) are increasingly of interest as one of the advanced techniques to mitigate the fading effects of transmission channel. We proposed a novel cooperative scheme with data exchange between relays before using distributed space-time coding. This fDSTC (full Distributed Space-Time Code) was compared with the conventional distributed space-time coded (cDSTC) protocol. Then, the thorough comparison of the fDSTC and cDSTC protocols in case of non-regenerative relays (NR-relays) and regenerative relays (R-relays) were considered in terms of error performance, outage probability, diversity order and energy consumption via both numerical simulations and mathematical analysis [24]. The previous works consider the energy efficiency of the cooperative relays techniques under the view of ideal medium access control (MAC) protocol. However, MAC protocol is responsible for regulating the shared wireless medium access of the networks, therefore, it has great influences on the total energy consumption of the networks. That lead us to a big motivation to design a cooperative MAC protocol, RIC-MAC (Receiver Initiated Cooperative MAC), by combining preamble sampling and cooperative relay techniques. The analytic results still confirm the interest of using cooperative relay techniques. However, the energy efficiency of the cooperative relay systems may be affected by MAC protocol design, the traffic loads of the networks and the desired latency [24].

6.3.6. Opportunistic Routing

Participants: Olivier Berder, Olivier Sentieys, Ruifeng Zhang.

However, the aforementioned approaches introduce an overhead in terms of information exchange, increasing the complexity of the receivers. A simpler way of exploiting spatial diversity is referred to as opportunistic routing. In this scheme, a cluster of nodes still serves as relay candidates but only a single node in the cluster forwards the packet [80]. Energy efficiency and transmission delay are very important parameters for wireless multihop networks. Numerous works that study energy efficiency and delay are based on the assumption of reliable links. However, the unreliability of channels is inevitable in wireless multihop networks. We investigated the tradeoff between the energy consumption and the latency of communications in a wireless multihop network using a realistic unreliable link model [43]. It provided a closed-form expression of the lower bound of the energy-delay tradeoff and of energy efficiency for different channel models (additive white Gaussian noise, Rayleigh fast fading and Rayleigh block-fading) in a linear network. These analytical results are also verified in 2-dimensional Poisson networks using simulations. The closed-form expression provides a framework to evaluate the energy-delay performance and to optimize the parameters in physical layer, MAC layer and routing layer from the viewpoint of cross-layer design during the planning phase of a network.

6.3.7. Adaptive Techniques for WSN Power Optimization

Participants: Olivier Berder, Daniel Menard, Olivier Sentieys, Mahtab Alam, Trong-Nhan Le.

We proposed a self-organized asynchronous medium access control (MAC) protocol for wireless body area sensor (WBASN). A body sensor network exhibits a wide range of traffic variations based on different physiological data emanating from the monitored patient. In this context, we exploit the traffic characteristics being observed at each sensor node and propose a novel technique for latency-energy optimization at the MAC layer [48], [26]. The protocol relies on dynamic adaptation of wake-up interval based on a traffic status register bank. The proposed technique allows the wake-up interval to converge to a steady state for variable traffic rates, which results in optimized energy consumption and reduced delay during the communication. The results show that our protocol outperforms the other protocols in terms of energy as well as latency under the variable traffic of WBASN.

System lifetime is the crucial problem of Wireless Sensor Networks (WSNs), and exploiting environmental energy provides a potential solution for this problem. When considering self-powered systems, the Power Manager (PM) plays an important role in energy harvesting WSNs. Instead of minimizing the consumption energy as in the case of battery powered systems, it makes the harvesting node converge to Energy Neutral Operation (ENO) to achieve a theoretically infinite lifetime and maximize the system performance. In [62], a low complexity PM with a Proportional Integral Derivative (PID) controller is introduced. This PM monitors the buffered energy in the storage device and performs adaptation by changing the wake-up period of the wireless node. This shows the interest of our approach since the impractical monitoring harvested energy as well as consumed energy is not required as it is the case in other previously proposed techniques. Experimental results are performed on a real WSN platform with two solar cells in an indoor environment. The PID controller provides a practical strategy for long-term operations of the node in various environmental conditions.

6.3.8. WSN for Health Monitoring

Participants: Patrice Quinton, Olivier Sentieys.

Applications of wireless sensor devices were also considered in the domain of health monitoring. Together with researchers from CASA team of IRISA-UBS, we investigated the possibility of using ECG-sensors to remotely monitor the cardiac activity of runners during a marathon race, using off-the shelf sensing devices and a limited number of base stations deployed along the marathon route. Preliminary experiments showed that such a scenario is indeed viable, although special attention must be paid to balancing the requirements of ECG monitoring with the constraints of episodic, low-rate transmissions.

The proliferation of private, corporate and community Wi-Fi hotspots in city centers and residential areas opens up new opportunities for the collection of biomedical data produced by sensors carried by mobile non-hospitalized subjects. Using disruption-tolerant networks, it was shown that biomedical data could be recorded using nearby hotspot. A scenario involving a subject wearing an ECG-enabled sensor walking in the streets of a residential area was reported.

These researches, combined with new sensor devices developed by the BOWI project, open up a large range of applications where high-performance sensor devices would allow health monitoring, or sport events organization.

6.3.9. Reconfigurable Video Coding

Participants: Emmanuel Casseau, Hervé Yviquel.

In the field of multimedia coding, standardization recommendations are always evolving. To reduce design time taking benefit of available SW and HW designs, Reconfigurable Video Coding (RVC) standard allows defining new codec algorithms. The application is represented by a network of interconnected components (so called actors) defined in a modular library and the behaviour of each actor is described in the specific RVC-CAL language. Dataflow programming, such as RVC applications, express explicit parallelism within an application. However general purpose processors cannot cope with both high performance and low power consumption requirements embedded systems have to face. Hence we are investigating the mapping of RVC specifications on hardware accelerators or on many tiny core platforms. Actually, our goal is to propose an automated co-design flow based on the Reconfigurable Video Coding framework. The designer provides the application description in the RVC-CAL dataflow language, after which the co-design flow automatically generates a network of processors that can be synthesized on FPGA platforms. We are currently focussing on a many-core platform based on the TTA processor (Very Long Instruction Word -style processor). Hervé Yviquel did a 4-months stay (Spring 2012) at Tampere University of Technology, Finland, in the group of Jarmo Takala who is developing a co-design toolset for TTA processor automated generation. Such a methodology permits the rapid design of a many-core signal processing system which can take advantage of all levels of parallelism. This work is done in collaboration with Mickael Raulet from IETR INSA Rennes and has been implemented in the Orcc open-source compiler. At present time the mapping of the RVC-CAL actor network is straightforward: every actor is mapped on a TTA processor based on our collaboration with Jani Boutellier from the University of Oulu (Finland). To reduce the area of the platform, TTA processor usage rate has to be improved, i.e. several actors are to be mapped onto a single processor. Work in progress is about this. It requires an actor partitioning step to define the set of actors that will be executed on the same processor. Due to the dynamic behaviour of the application, we expect we will be able to use profiling to get some feedbacks for the partitioning.

6.3.10. A Low-Complexity Synchronization Method for OFDM Systems

Participants: Pramod P. Udupa, Olivier Sentieys, Pascal Scalart.

A new hierarchical synchronization method was proposed for initial timing synchronization in orthogonal frequency-division multiplexing (OFDM) systems. Based on the proposal of new training symbol, a threshold based timing metric is designed for accurate estimation of start of OFDM symbol in a frequency selective channel. Threshold is defined in terms of noise distributions and false alarm which makes it applicable independent of type of channel it is applied. Frequency offset estimation is also done for the proposed training symbol. The performance of the proposed timing metric is evaluated using simulation results. The proposed method achieves low mean squared error (MSE) in timing offset estimation at five times lower computational complexity compared to cross-correlation based method in a frequency selective channel. It is also computationally efficient compared to hybrid approaches for OFDM timing synchronization.

6.3.11. Flexible hardware accelerators for biocomputing applications

Participants: Steven Derrien, Naeem Abbas, Patrice Quinton.

It is widely acknowledged that FPGA-based hardware acceleration of compute intensive bioinformatics applications can be a viable alternative to cluster (or grid) based approach as they offer very interesting MIPS/watt figure of merits. One of the issues with this technology is that it remains somewhat difficult to use and to maintain (one is rather designing a circuit rather than programming a machine). Even though there exists C-to-hardware compilation tools (Catapult-C, Impulse-C, etc.), a common belief is that they do not generally offer good enough performance to justify the use of such reconfigurable technology. As a matter of fact, successful hardware implementations of bio-computing algorithms are manually designed at RT-level and are usually targeted to a specific system, with little if any performance portability among reconfigurable platforms. This research work, funded by the ANR BioWic project, aims at providing a framework for helping semi-automatic generation of high-performance hardware accelerators. This research work builds upon the CAIRN research group expertise on automatic parallelization for application specific hardware accelerators and has been targeting mainstream bioinformatics applications (HMMER, ClustalW and BLAST). The Biowic project ended in early 2012. Naeems Abbas, a PhD student funded by the project defended his PhD in May 2012.

CAMUS Team

6. New Results

6.1. VMAD

Participants: Alexandra Jimborean, Philippe Clauss, Jean-François Dollinger, Aravind Sukumaran-Rajam, Juan Manuel Martinez Caamaño, Vincent Loechner.

The goal of the VMAD project is to provide a set of annotations (pragmas) that the user can insert in the source code to perform advanced analyses and optimizations, for example dynamic speculative parallelization.

VMAD contains a modified LLVM compiler and a runtime system. The program binary files are first generated by our compiler to include necessary data, instrumentation instructions, parallel code templates, and callbacks to the runtime system. External modules associated to specific analyses and transformations are dynamically loaded when required at runtime. Dynamic information, such as memory locations of the modules entries, are patched at startup in the loaded executable.

VMAD uses sampling and multi-versioning to limit the runtime overhead (profiling, analysis, and code generation). At runtime, targeted codes are launched by successive chunks that can be either original, instrumented or optimized/parallelized versions. After each chunk execution, decisions can be taken relatively to the current optimization strategy. VMAD is handling advanced memory access profiling [17] through linear interpolation of the addresses, dynamic dependence analysis, version selection [17] and speculative polyhedral parallelization [19], [16].

Alexandra Jimborean defended her PhD thesis on this topic in 2012 [12]. In 2012, Aravind Sukumaran-Rajam started a PhD in our team to continue this work, especially on extending the dependence analysis to make it handle more general programs, keeping it fast and accurate. Jean-François Dollinger will extend the framework to handle heterogeneous architectures (GPGPUs). Juan Manuel Martinez Caamaño, a master student of University of Buenos Aires (associate team EA-Ancome) is also working on VMAD to make the code generation support tiling.

6.2. The Multifor programming construct

Participants: Philippe Clauss, Imèn Fassi, Yosr Slama, Matthieu Kuhn.

We have proposed a new programming control structure called “multifor”, allowing to take advantage of parallelization models that were not naturally attainable with the polytope model before. In a multifor-loop, several loops whose bodies are run simultaneously can be defined. Respective iteration domains are mapped onto each other according to a run frequency – the grain – and a relative position – the offset –. Execution models like dataflow, stencil computations or MapReduce can be represented onto one referential iteration domain, while still exhibiting traditional polyhedral code analysis and transformation opportunities. Moreover, this construct provides ways to naturally exploit hybrid parallelization models, thus significantly improving parallelization opportunities in the multicore era. Traditional polyhedral software tools are used to generate the corresponding code. Additionally, a promising perspective related to non-linear mapping of iteration spaces has also been developed, yielding to run a loop nest inside any other one by solving the problem of inverting “ranking Ehrhart polynomials”.

This work is the PhD work of Imèn Fassi, who started her work this year and who is co-advised by Yosr Slama, Assistant Professor at the University El Manar in Tunis, Tunisia, and Philippe Clauss. A first publication of this topic has been accepted at the IMPACT workshop that will be held in conjunction with the HIPEAC conference in Berlin, Germany, January 2013.

6.3. Parwiz: dynamic data dependence analysis

Participants: Alain Ketterlin, Philippe Clauss.

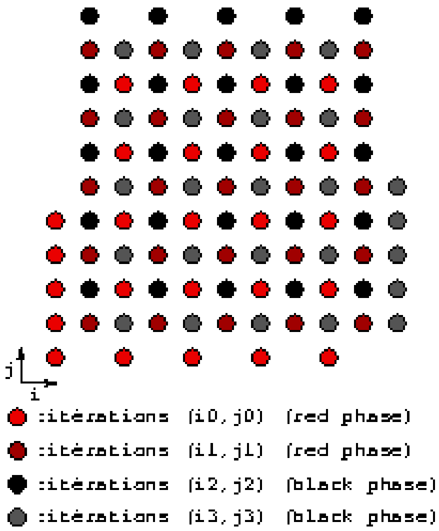


Figure 3. Red-Black Gauss-Seidel Multifor Iteration Space

We have continued working on dynamic data-dependence analysis during this year, especially on increasing the scope of our tool (called Parwiz). For instance, Parwiz is now able to suggest several program transformations (like loop distribution) that enable loop vectorization. It uses an algorithm known as *codegen* (developed by Allen & Kennedy), but the novelty is that it applies the algorithm to dependence graphs that are built empirically, by running the program on selected input data sets. As far as we know, Parwiz is the first tool able to suggest loop transformations.

We have also developed several other empirical analysis. One of these focuses on loops that are not parallel, but whose iterations present significant parallelism provided the program explicitly schedules the various iterations. This still lacks a suitable cost model to estimate the potential gain, but gives significant insight into the behavior of a given non-parallel loop.

This work has been presented at the MICRO-45 conference held in Vancouver on december 1–5 2012 [18].

6.4. Modeling the behavior of parallel traces

Participants: Alain Ketterlin, Stéphane Genaud.

We have started this year a project aiming at developing algorithms and tools to capture the behavior of parallel programs. Our initial goal is automatically obtain formal models of communicating MPI processes, in terms of message sends and receives and of synchronization events. Such models have various uses, the first of them being the visualization of the system's communications, for debugging, or plain understanding (see below, Figure 4). However, we expect to develop other applications, for example in optimizing the communication infrastructure (or routing algorithm) for specific applications.

Our modeling algorithm works in two phases. The first phase is local to each node, using our work on nested loop recognition [7]. This builds a sequence of loop nests providing a compact representation of all local communication events. At the end of the run, the various local models are merged, typically through a parallel reduction operation, to build the global model.

We plan to publish the first part of this work in the first half of 2013. Several experimental data have been collected already, but we would like to evaluate the overall task on significantly sized programs.

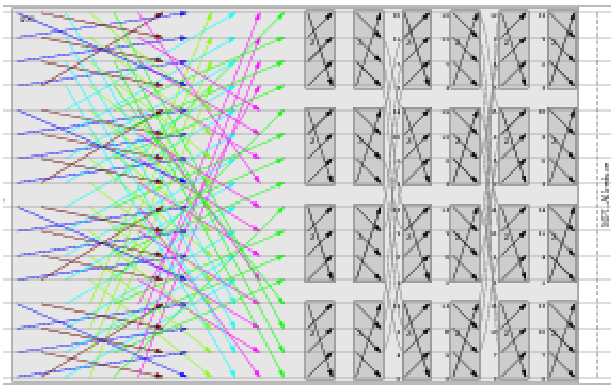


Figure 4. Visualizing parallel traces

Currently, the whole process is restricted to communication events. However, it can be immediately extended to trace including other kinds of events, like the addresses and sizes of memory buffers transmitted from process to process. This would provide a complete, run time description of the program, which could be used to evaluate the potential gain of various re-parallelization techniques. This aspect is the next goal on our agenda.

6.5. Certified polyhedral transformations into more and more concrete languages

Participants: Nicolas Magaud, Julien Narboux, Éric Violard.

We continued our work to complete the proof of polyhedral based transformations in the language *Loops* designed by Alexandre Pilkiewicz (see the proof scheme on Fig. 5). Our idea is to use once again a validator. The validation here consists in comparing two polyhedrons: the one (**pprogopt**) obtained from the non-optimized Loops program (**prog**), by translation to the polyhedral language (*Plang*) (**pprog**), and then optimization in *Plang*; and the one (**interprogopt**) obtained from the optimized Loops program (**progopt**) by translation into *Plang*. If these two polyhedrons are the same, then the validator returns true, otherwise it returns false. The proof that the non-optimized and optimized programs have the same behaviour lies on the deterministic property of the function that translates a program *Loops* into *Plang*. We obtained the proof in Coq that our scheme is correct. Now, we have to complete the implementation of our optimizing compiler for *Loops* by connecting our validator with the off the shell tools for polyhedral transformations. We will use the tool PLuTo[10] to find efficient code transformations and CLooG[11] to generate the loops from the polyhedral representation (we proposed an internship for this purpose).

We now have to connect the language *Loops* with more concrete languages (whose features and semantics have to be defined). We already showed how to deal with arithmetic overflows in a more concrete language where each loop variable is a machine integer [20]. Our approach is thus to incrementally add concrete features until joining an intermediate language of CompCert.

Since the members of our team have some skill in defining new languages and their semantics, we thought that it could be a good idea to exploit this and to define a formal semantics for the **Multifor** syntactic sugar proposed by Philippe Clauss. We aims at associating a rigorous mathematical meaning with this syntactic construct: first a denotational semantics and then an operational one. This work will serve as a base to prove correct the compilation process that translates this construct into intermediate code.

[10] http://pluto-compiler.sourceforge.net/
[11] http://www.cloog.org/

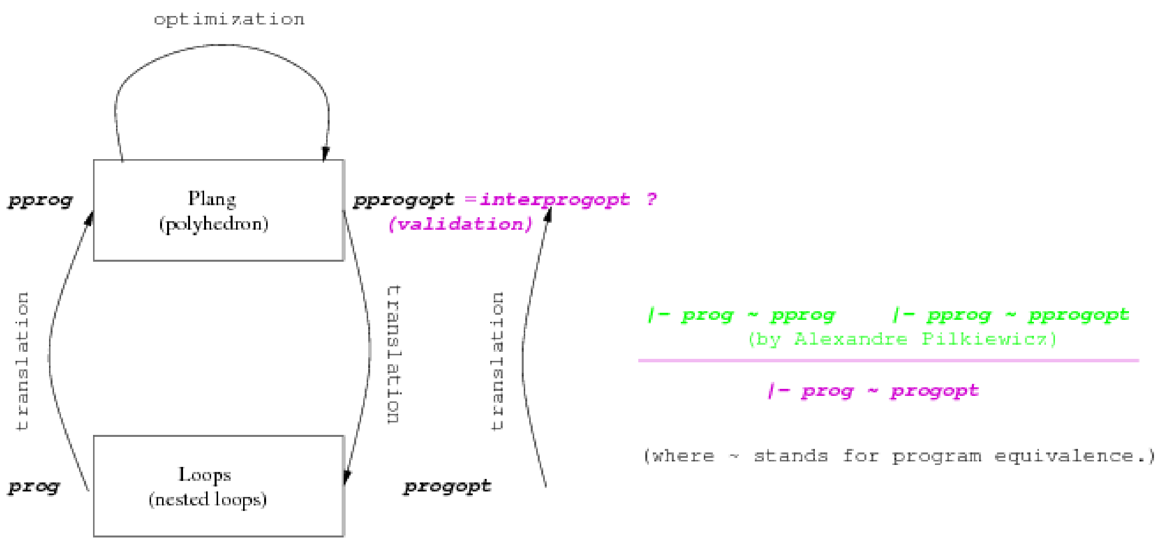


Figure 5. Our proof scheme for a certified compiler of Loops

CARAMEL Project-Team

6. New Results

6.1. Sieve for FFS

Participants: Jérémie Detrey, Pierrick Gaudry [contact], Marion Videau.

Jérémie Detrey, Pierrick Gaudry and Marion Videau have worked on the relation collection step of the Function Field Sieve and especially on its implementation. This is still an ongoing work but the first results have been accepted for publication [13] in the ARITH-2013 conference.

6.2. Bilinear Maps

Participants: Răzvan Bărbulescu, Jérémie Detrey, Nicolas Estibals, Paul Zimmermann [contact].

As a result of an internal working group in the team, we have found and published at the WAIFI conference a new algorithm to find optimal formulae for bilinear maps [8]. This algorithm enables one to rediscover Karatsuba's multiplication algorithm, but has many other applications, for example to matrix multiplication.

6.3. Number Field Sieve

Participants: Emmanuel Thomé, Paul Zimmermann [contact].

Together with Shi Bai (Australian National University), E. Thomé and P. Zimmermann used CADO-NFS to factor RSA-704, a 212-digit number, to check scalability of the software on large factorizations [10]. This is the second largest number factored by any GNFS software so far, and the largest one factored by CADO-NFS. This experiment was very helpful, since it demonstrated several weaknesses of the code, that have been addressed since then.

Together with Shi Bai (Australian National University), P. Zimmermann wrote a preprint describing the algorithm used in CADO-NFS for the size-optimization of sextic polynomials [11].

Alain Filbois, Shi Bai (Australian National University) and P. Zimmermann improved the polynomial selection code. With parameters used to find good polynomials for RSA-896, a total speedup by a factor 14 was obtained, with both algorithmic and implementation improvements.

6.4. Sparse linear algebra modulo p

Participants: Hamza Jeljeli, Emmanuel Thomé [contact].

The resolution of linear algebra problems with subexponential methods, which is the topic of the ANR-CATREL project (to begin in 2013) calls for the resolution of large sparse linear systems defined over finite fields. In preparation for this, H. Jeljeli has developed software for performing sparse matrix times vector multiplication on NVIDIA GPUS [16]. This code provides a very significant speedup over the use of CPUs for this task, and achieves this speedup by a clever use of a "residue number system" representation of the finite field elements.

As a complement, a recent re-implementation of Thomé's algorithm for the (matrix) Berlekamp-Massey step in the block Wiedemann algorithm has been done. This program can of course be special-cased to the simple non-matrix case. The GPU code above and this special case, together, form the needed software to have a sparse linear system solver over finite fields using Wiedemann's algorithm. This has been put to use, and led to the completion of a discrete logarithm record in $\mathbb{F}_{2^{619}}$, the linear system part taking only 17 hours in total on one GPU (plus 1 hour on one CPU for the Berlekamp-Massey step).

6.5. Using symmetries in elliptic curve discrete logarithm

Participant: Pierrick Gaudry.

In a joint work by Jean-Charles Faugère, Pierrick Gaudry, Louise Huot and Guénaël Renault, it has been shown that the geometric symmetries of an elliptic curve, in particular, the symmetries of an Edwards curve, could be used to speed up the index calculus attack for computing discrete logarithms in an elliptic curve defined over an extension field. The corresponding article [14] is currently under revision.

6.6. Galois properties of curves for ECM

Participants: Răzvan Bărbulescu, Cyril Bouvier.

In collaboration with Joppe Bos, Peter Montgomery and Thorsten Kleinjung, Răzvan Bărbulescu and Cyril Bouvier proved some divisibility properties of the group order of an elliptic curve, using the Galois structure of its division polynomial. It explains the good behaviour of some curves that have been experimentally found to factor more numbers than others, and gives a way to find new curves with this property. The corresponding article [7] was presented in ANTS-X.

6.7. Computation of CM class polynomials for genus 2 Jacobians

Participants: Sorina Ionica, Emmanuel Thomé [contact].

In collaboration with Andreas Enge, Emmanuel Thomé has developed software for computing class polynomials, in the context of complex multiplication theory in genus 2. The current computations set new records which are well above the previous state of the art. A publication is in the works.

Using similar underlying tools and theory, and based on work by Sorina Ionica [15], Sorina Ionica and Emmanuel Thomé have worked on the analysis of isogeny graphs in genus 2, when certain properties of the endomorphism ring are satisfied.

6.8. Filtering step for NFS and FFS

Participant: Cyril Bouvier.

Cyril Bouvier studied the filtering step for the Number Field Sieve. A better weight function, used during the clique removal step, was found which allows to construct smaller matrices for the linear algebra step. A preprint is avalaible [12]. The filtering step for the Function Field Sieve was written in CADO-NFS.

CARTE Project-Team

6. New Results

6.1. Dynamical systems

Participant: Mathieu Hoyrup.

Birkhoff theorem is a central result in ergodic theory. Consider a dynamical system $(X, T : X \to X)$, start with an initial condition $x \in X$ and construct the trajectory $(x, T(x), T^2(x), ...)$. How is this trajectory distributed in X? What is the limit frequency of visits of a set $A \subseteq X$? Ergodic theorems answer to these questions by showing (i) that the distribution of *almost every* point converges and (ii) by describing the possible distributions associated to trajectories.

For several years we have been working on the project of identifying the exact computational content of several ergodic theorems: can the speed of convergence of limit frequencies be computed? Can one distinguish between points with different limit frequencies? Can we construct (compute) points whose trajectory follow a prescribed distribution? How random (i.e. incompressible) a point has to be for the distribution of its trajectory to converge?

6.1.1. Limit frequencies

We have obtained new insight in the above questions by proving that random elements eventually reach effective closed sets of positive measure (while it was only known for a more restricted class of sets). The paper appeared in *Information and Computation* [11]. This result is a key tool in the proof of the result published in [23].

6.1.2. Information

A chaotic system is unpredictable because it has much more trajectories than observable initial conditions: hence many undistinguishable initial points lead to radically different trajectories. As there are many trajectories, most of them are complex in the sense that they can hardly be compressed, i.e. described in a shorter way than simply listing them. The Shannon-McMilan-Breiman theorem states that the compression-rate of most trajectories coincides with the entropy of the system.

We have been interested in the computational content of this theorem: how random a point has to be to generate a trajectory whose compression rate is the entropy? This question was raised in [71] and has been left open for 14 years. We have solved the problem by showing that Martin-Löf notion of randomness is sufficient. Our recent result presented in [11]is a key ingredient of our proof. We presented the result at *STACS* [23].

6.1.3. Decomposition

The ergodic decomposition theorem states that a dynamical system can always be uniquely decomposed into indecomposable subsystems, technically *ergodic* subsystems. We have been interested in the computability of the decomposition operation. It is known from [71] that this operation is not computable in general. Whether this operation is still not computable when the system can be decomposed into a *finite* number of subsystems was open. We raised the question and answer it negatively in [57]. More precisely, we prove the existence of ergodic measures P and Q such that neither P nor Q is computable relative to $P + Q$. In other words, the operation of splitting a non-ergodic process into ergodic components is not computable, even in the trivial case of a combination of 2 ergodic processes. The paper is currently in press and will appear in *Annals of Pure and Applied Logic* [14].

6.2. Computations

Participant: Mathieu Hoyrup.

6.2.1. Inversion of computable functions

We strengthen the preceding result by making P and Q computable. This result is a particular case of a more general problem. In many situations an operator $F \to Y$ can be computed but can hardly be reversed: given $F(x)$, x cannot always we recovered (computed) even when F is one-to-one. We introduce a strong notion of discontinuity for the inverse of F and prove that it entails the existence of a non-computable $x \in X$ such that $F(x)$ is computable. Our result on the ergodic decomposition can be derived by applying our general result to the operator $F(P,Q) = P + Q$ which is computable but difficult to reverse. At the same time we prove a significant improvement of a classical result of Pour-El and Richards [67] about the computability of linear operators. The paper [26] is currently submitted.

6.2.2. Computability and measure theory.

We study the constructive content of the Radon-Nikodym theorem, show that it is not computable in general and precisely locate its non-computability in the Weihrauch lattice. The paper [15] appeared in the first issue of the new journal *Computability*.

6.3. Computer virology

6.3.1. Behavioral analysis

Participants: Isabelle Gnaedig, Jean-Yves Marion.

Our study on behavioural malware detection has been continued. We have been developing an approach detecting suspicious schemes on an abstract representation of the behavior of a program, by abstracting program traces, rewriting given subtraces into abstract symbols representing their functionality. Considering abstract behaviors allows us to be implementation-independent and robust to variants and mutations of malware. Suspicious behaviors are then detected by comparing trace abstractions to reference malicious behaviors.

We had previously proposed to abstract trace automata by rewriting them with respect to a set of predefined behavior patterns defined as a regular language described by a string rewriting system [32]. We then have increased the power of our approach on two aspects. We fist have modified the abstraction mechanism, keeping the abstracted patterns in the rewritten traces, which allows us to handle interleaved patterns. Second, we have extended the rewriting framework to express data constraints on action parameters by using term rewriting systems. An important consequence is that, unlike in [32], using the data-flow, we can now detect information leaks in order to prevent unauthorized disclosure or modifications of information.

We also have introduced model checking in our approach: the predefined behavior patterns, used to abstract program traces, have been defined by first order temporal logic formulas, as well as the reference suspicious behaviors, given in a signature. The infection problem can then be seen as the satisfaction problem of the formula of the signature by an abstracted trace of the program, which can be checked using existing model checking techniques. This work has been published at the ESORICS conference [20].

6.3.2. Analyzing cryptographic implementations

Participants: Joan Calvet, Jean-Yves Marion.

Analyzing cryptographic implementations has important applications, especially for malware analysis where they are an integral part both of the malware payload and the unpacking code that decrypts this payload. These implementations are often based on well-known cryptographic functions, whose description is publicly available. While potentially very useful for malware analysis, the identification of such cryptographic primitives is made difficult by the fact that they are usually obfuscated. Current state-of-the-art identification tools are ineffective due to the absence of easily identifiable static features in obfuscated code. However, these implementations still maintain the input-output (I/O) relationship of the original function. In a joint work with José M. Fernandez published in [22], we present a tool that leverages this fact to identify cryptographic functions in obfuscated programs, by retrieving their I/O parameters in an implementation-independent fashion, and comparing them with those of known cryptographic functions. In experimental evaluation, we successfully identified the cryptographic functions TEA, RC4, AES and MD5 in obfuscated programs. In addition, our tool was able to recognize basic operations done in asymmetric ciphers such as RSA.

6.3.3. Self-replication

Participant: Jean-Yves Marion.

Self-replication is one of the fundamental aspects of computing where a program or a system may duplicate, evolve and mutate. Our point of view is that Kleene's (second) recursion theorem is essential to understand self-replication mechanisms. An interesting example of self-replication codes is given by computer viruses. This was initially explained in the seminal works of Cohen and of Adleman in the eighties. In fact, the different variants of recursion theorems provide and explain constructions of self-replicating codes and, as a result, of various classes of malware. None of the results are new from the point of view of computability theory. We just propose a self-modifying register machine as a model of computation in which we can effectively deal with self-reproduction and in which new offsprings can be activated as independent organisms. This work was published by Jean-Yves Marion in a special issue on the honor of Alan Turing [16].

6.3.4. Reverse engineering by morphological analysis

Participants: Guillaume Bonfante, Jean-Yves Marion, Fabrice Sabatier, Aurélien Thierry.

Let us suppose we are given some malware and we want to know what it is doing. One may run it, or one may analyze it more or less statically. Typically, an expert tries to guess the behavior of a malware through the analysis of its binary code (in tools such as Ida). The task is much simpler if the expert already knows some part of the code. We have shown that morphological analysis could be used in such a context. We have rediscovered the parts of the malware `Duqu` within `Stuxnet`. We have rediscovered the compilation options used to include OpenSSL's functions within `Waledac` [21].

CASCADE Project-Team (section vide)

CASSIS Project-Team

6. New Results

6.1. Automated Deduction

We develop general techniques which allow us to re-use available tools in order to build a new generation of solvers offering a good trade-off between expressiveness, flexibility, and scalability. We focus on the careful integration of combination techniques and rewriting techniques to design decision procedures for a wide range of verification problems.

6.1.1. Building and verifying decision procedures

Participants: Alain Giorgetti, Olga Kouchnarenko, Christophe Ringeissen, Elena Tushkanova.

We have developed a methodology to build decision procedures by using superposition calculi which are at the core of equational theorem provers. We are interested in developing automated deduction techniques to prove properties about these superposition-based decision procedures. To this aim, we plan to further investigate the use of schematic superposition, which has been already applied to check the termination and the combinability of superposition-based procedures. We have been working on the development of a framework for specifying and verifying superposition-based procedures. In [52], we present an implementation in Maude of the two inference systems corresponding to superposition and schematic superposition. Thanks to this implementation we automatically derive termination of superposition for a couple of theories of interest in verification.

Until now, schematic superposition was only studied for standard superposition. In [62], we introduce a schematic superposition calculus modulo a fragment of arithmetics, namely the theory of Integer Offsets. This new schematic calculus is used to prove the decidability of the satisfiability problem for some theories extending Integer Offsets. We illustrate our theoretical contribution on theories representing extensions of classical data structures, e.g., lists and records. Our Maude-based implementation has been extended to incorporate this new schematic superposition calculus modulo Integer Offsets. It enables automatic decidability proofs for theories of practical use.

6.2. Security Protocol Verification

The design of cryptographic protocols is error-prone. Without a careful analysis, subtle flaws may be discovered several years after the publication of a protocol, yielding potential harmful attacks. In this context, formal methods have proved their interest for obtaining good security guarantees. Many analysis techniques have been proposed in the literature [72]. We have edited a book [65] where each chapter presents an important and now standard analysis technique. We develop new techniques for richer primitives, wider classes of protocols and higher security guarantees.

6.2.1. Equational theories of cryptographic primitives

Participant: Michaël Rusinowitch.

Some attacks exploit in a clever way the interaction between protocol rules and algebraic properties of cryptographic operators. In [76], we provide a list of such properties and attacks as well as existing formal approaches for analyzing cryptographic protocols under algebraic properties.

Encryption "distributing over pairs" is employed in several cryptographic protocols. We have shown that unification is decidable for an equational theory HE specifying such an encryption [15]. We model block chaining in terms of a simple, convergent, rewrite system over a signature with two disjoint sorts: list and element and present in [27] an algorithm for deciding the unification problem modulo this rewrite system. Potential applications of this unification procedure include flaw detection for protocols employing the CBC encryption mode. We have also studied a very simple property satisfied by the RSA-based implementation of the *blind signature scheme* and we have shown its unification problem is undecidable [28]. It is the simplest theory, to our knowledge, for which unification is undecidable.

In their seminal work Dolev and Yao used string rewriting to check protocol security against an active intruder. The main technical result and algorithm were improved by Book and Otto who formulated the security check in terms of an extended word problem for cancellation rules. We extend in [16] their main decidability result to a larger class of string rewrite systems called opt-monadic systems.

6.2.2. Voting protocols

Participants: Mathilde Arnaud, Véronique Cortier, David Galindo-Chacon, Stéphane Glondu, Malika Izabachene, Steve Kremer, Cyrille Wiedling.

Voting is a cornerstone of democracy and many voting systems have been proposed so far, from old paper ballot systems to purely electronic voting schemes. Although many works have been dedicated to standard protocols, very few address the challenging class of voting protocols. We have studied several protocols that are currently in use:

- Helios is an open-source web-based end-to-end verifiable electronic voting system, used e.g. by UCL and the IACR association in real elections. We have discovered a vulnerability which allows an adversary to compromise the privacy of voters and we have presented a fixed version, showed to satisfy a formal definition of ballot secrecy using the applied pi calculus [21]. One main advantage of Helios is its verifiability, up-to the ballot box (a dishonest ballot box may add ballots). We are now working on defining a variant of Helios that prevents from ballot stuffing, even against a dishonest ballot box. Our approach consists in introducing an additional authorities that provides credentials that the ballot box can verify but not forge. This new version is under implementation and we are proving computational security for both ballot secrecy (inherited from Helios) and full verifiability (due to our credentials).
- Norway has used e-voting in its last political election in September 2011, with more than 25 000 voters using the e-voting option. Using formal models, we have analyzed the underlying protocol w.r.t. privacy, considering several corruption scenarios [41].
- The Section 07 of CNRS (now split into Section 06 and Section 07) has proposed a voting protocol for Face-to-Face meetings to enhanced the verifiability of an election run through electronic devices. We have formally modeled this protocol and proved both ballot secrecy and verifiability.

Even a basic property like ballot secrecy is difficult to define formally and several definitions co-exist. The loss of privacy may not only come from the protocol but also from the tally function itself and depends on what needs to be kept private. We have proposed a general and quantitative definition of privacy, that captures two previously proposed definitions [35]. Security based on cryptography relies on the fact that certain operations (such as decrypting) are computationally infeasible. However, e-voting protocols should also guarantee privacy in the future, when computers will have an increased computational power and will be able e.g. to break nowadays keys. Such privacy in the future is called *everlasting privacy* and we have proposed a definition of *practical everlasting privacy*.

6.2.3. Other families of protocols

Participants: Véronique Cortier, Steve Kremer, Robert Künnemann, Cyrille Wiedling.

Securing routing Protocols. The goal of routing protocols is to construct valid routes between distant nodes in the network. If no security is used, it is possible for an attacker to disorganize the network by maliciously interacting with the routing protocols, yielding invalid routes to be built. That is why secure versions of routing protocols are now developed. The security model differs from standard protocols since the adversary can only control some nodes of the network. The security of a routing protocols therefore depends on the network topology. In [39], we show a simple reduction result: if there is an attack then there is an attack in a four nodes topology. It is therefore sufficient to study security for a finite number of distinct topologies, allowing to reuse existing tools such as ProVerif.

Security APIs. In some systems, it is not possible to trust the host machine on which sensitive codes are executed. In that case, security-critical fragments of a program should be executed on some tamper resistant device (TRD), such as a smartcard, USB security token or hardware security module (HSM). The exchanges between the trusted and the untrusted infrastructures are ensured by special kind of API (Application Programming Interface), that are called *security APIs*. We have previously designed a generic API for key-management based on key hierarchy [77]. In [40], [60], we have extended our API to handle key-revocation such that the security tokens can still be used (it is not necessary to revoke the full token) and such that any key can be revoked (even upper keys in the hierarchy). In [64], we propose a universally composable key management functionality and show how to achieve a secure, distributed implementation on TRDs.

6.2.4. Automated verification of indistinguishability properties.

Participants: Rémy Chrétien, Véronique Cortier, Steve Kremer.

New emerging classes of protocols such as voting protocols often require to model less classical security properties, such as anonymity properties, strong versions of confidentiality and resistance to offline guessing attacks. Many of these properties can be modelled using the notion of indistinguishability by an adversary, which can be conveniently modeled using process equivalences.

Static case. The YAPA tool [17] can check static equivalence for convergent equational theories. It is proved to terminate for a wide class of equational theories that includes subterm convergent theories (e.g. encryption, signatures, pairing and hash) and layered convergent theories (e.g. blind signatures). The procedure is generic in the sense that it remains sound and complete (but may not terminate) for any convergent theory. It has been implemented in the YAPA tool[3]. The KISS tool [19] is also able to verify static equivalence for convergent equational theories. Termination has been shown for subterm convergent equational theories (a subset of layered convergent theories) as well as several equational theories motivated by electronic voting protocols such as blind signatures and trap-door commitment schemes (which are out of the scope of YAPA).

In [20], we show how to *combine* decision procedures: if static equivalence and deduction are decidable for two disjoint equational theories then they are decidable for the union of the theories. In [25] we develop a method that allows us in some cases to simplify the task of deciding static equivalence in a multi-sorted setting, by removing a symbol from the term signature and reducing the problem to several simpler equational theories. We illustrate our technique at hand of bilinear pairings.

Active case. In [36] we present a novel procedure to verify equivalence properties for a bounded number of sessions which is able to handle a large class of equational theories. Although, we were unable to prove termination of the resolution procedure, the procedure has been implemented in a prototype tool and has been effectively tested on examples. We were able to verify properties such as guessing attacks in password protocols, strong flavors of confidentiality and anonymity properties, including fully automated checking of anonymity of an electronic voting protocol by Fujioka et al. which was outside the scope of existing tools.

In [42] we study this equivalence problem when cryptographic primitives are modeled using a group equational theory, a special case of monoidal equational theories. We reduce the problem to solving systems of equations over rings and provide several new decidability and complexity results, notably for equational theories which have applications in security protocols, such as exclusive or and Abelian groups which may additionally admit a unary, homomorphic symbol.

Rémy Chrétien has recently started a PhD on deciding trace equivalence for an unbounded number of sessions. His first findings show that for some classes of protocols, decidability of trace equivalence can be reduced to equivalence of deterministic pushdown automata (which is decidable [81]).

Note that for simple processes without branch nor replication observational equivalence can be reduced to checking whether two symbolic constraints (representing honest agents) are equivalent [75]. We have published a new proof that symbolic constraints equivalence is decidable for the large class of subterm convergent theories [18].

[3] http://www.lsv.ens-cachan.fr/~baudet/yapa/

6.2.5. Soundness of the Dolev-Yao Model

Participants: Véronique Cortier, Guillaume Scerri.

All the previous results rely on symbolic models of protocol executions in which cryptographic primitives are abstracted by symbolic expressions. This approach enables significantly simple and often automated proofs. However, the guarantees that it offers have been quite unclear compared to cryptographic models that consider issues of complexity and probability. A recent line of research consists in identifying cases where it is possible to obtain the best of both cryptographic and formal worlds: fully automated proofs and strong, clear security guarantees.

Existing soundness results for symmetric encryption are not satisfactory. This is due to the fact that dishonest keys may introduce many behaviors that cannot be easily captured in symbolic models. Guillaume Scerri has started a PhD thesis on designing more flexible symbolic models for cryptographic proofs. His first result is a computationally sound symbolic model in the presence of dishonestly generated keys, allowing a symbolic adversary to generate new equalities between terms, on-the-fly [38].

6.3. Model-based Verification

We have investigated extensions of regular model-checking to new classes of rewrite relations on trees. We have studied specification and proof of modular imperative programs.

6.3.1. Algorithms for Tree Walking Automata

Participants: Pierre-Cyrille Héam, Vincent Hugot, Olga Kouchnarenko.

Tree walking automata are widely used to tackle data base algorithmic problems, particularly to analyse queries over XML documents. The emptiness problem for tree walking automata is known to be EXPTIME-complete. The general algorithm to solve this problem consists in transforming the tree walking automaton into a classical top-down tree automaton. The best known in the literature algorithm works in time $O(s2^{n^2})$ where n is the number of states of the tree walking automaton, and s is the size of the alphabet. In [24] we have proposed a new algorithm based on an *overloop* concept and working in time $O(2^{n^2})$. Then our approach has been improved for deterministic tree walking automata to have in this case a $O(2^{n \log n})$ time complexity. Finally, we have also proposed a polynomial-time approximation based semi-algorithm for the emptiness problem. The algorithms have been implemented and experimental results confirm the relevance of the approach.

6.3.2. Algorithms for Tree Automata with Global Constraints

Participants: Pierre-Cyrille Héam, Vincent Hugot, Olga Kouchnarenko.

Extending tree automata models to be able to compare different tree branches is an important and challening issue for systems' modeling and for verifying their properties. Several exetensions have been proposed in the litterature. Among them we are interested in the model of Tree Automata with Global Constraints (TAGED) introduced in 2009. The membership problem for this new model is known to be NP-complete, and the emptyness problem is known to be EXPTIME-complete. In [47] we have investigated some complexity results for tree automata with a bounded number of equality constraints. We have proved that with a unique constraint the emptyness problem is in PTIME and that it is EXPTIME-complete with only two constraints. For a bounded number of constraints, the membership problem is in PTIME.

6.3.3. Verification of Linear Temporal Patterns over Finite and Infinite Traces

Participants: Pierre-Cyrille Héam, Vincent Hugot, Olga Kouchnarenko.

In the regular model-checking framework, reachability analysis can be guided by temporal logic properties, for instance to achieve the counter example guided abstraction refinement (CEGAR) objectives. A way to perform this analysis is to translate a temporal logic formula expressed on maximal rewriting words into a "rewrite proposition" – a propositional formula whose atoms are language comparisons, and then to generate semi-decision procedures based on (approximations of) the rewrite proposition. In [46] we have investigated suitable semantics for LTL on maximal rewriting words and their influence on the feasibility of a translation, and we have proposed a general scheme providing exact results for a fragment of LTL corresponding mainly to safety formulæ, and approximations for a larger fragment.

6.3.4. Rewriting-based Mathematical Model Transformations

Participants: Walid Belkhir, Alain Giorgetti.

We have pursued our collaboration with the Department "Temps-Fréquence" of the FEMTO-ST institute (Franche-Comté Electronique Mécanique Thermique et Optique - Sciences et Technologies, CNRS UMR 6174) on the formalization of asymptotic methods (based on two-scale convergence) to automatically generate asymptotic models of large arrays of micro- and nanosystems. The goal is to provide engineers with an implementation of this mathematical tool inside a modeling software. We follow therefore a multidisciplinary approach which combines a generalization and formalization effort of mathematical asymptotic methods, together with rewriting-based formal transformation techniques. This approach is described in [53], together with an example and a presentation of the architecture of the software under design. A second contribution [34] is a detailed formal specification and analysis of lazy pattern-matching mechanism modulo associativity and commutativity, and its integration into a strategy language. The pattern-matching solutions are stored in a lazy list composed of a first substitution at the head and a non-evaluated object that encodes the remaining computations. Rule and strategy applications also produce a lazy list of terms. This contribution has been published in EPTCS as the proceedings of the 10th International Workshop on Reduction Strategies in Rewriting and Programming, where a lighter version was presented in 2011 [69].

6.4. Model-based Testing

Our research in Model-Based Testing (MBT) aims to extend the coverage of tests. The coverage refers to several artefacts: model, test scenario/property, and code of the program under test. The test generation uses various underlying techniques such as symbolic animation of models [22] or symbolic execution of programs by means of dedicated constraints, SMT solvers, or model-checkers.

6.4.1. Automated Test Generation from Behavioral Models

Participants: Fabrice Bouquet, Kalou Cabrera, Jérome Cantenot, Frédéric Dadeau, Elizabeta Fourneret, Jean-Marie Gauthier, Jonathan Lasalle.

We have introduced an original model-based testing approach that takes a behavioural view (modelled in UML) of the system under testing and automatically generates test cases and executable test scripts according to model coverage criteria. We continue to extended this result to SysML specifications for validating embedded systems [26]. To allow the test generation from SysML model, we study the transformation into a low level language more close of hardware in [44].

In the context of software evolution, we have worked on exploiting the evolution of requirements in order to classify test sequences, and precisely target the parts of the system impacted by this evolution. We have proposed to define the life cycle of a test via three test classes: (i) Regression, used to validate that unimpacted parts of the system did not change, (ii) Evolution, used to validate that impacted parts of the system correctly evolved, and (iii) Stagnation, used to validate that impacted parts of the system did actually evolve. The associated algorithms are under implementation in a dedicated prototype to be used in the SecureChange european project. A link with the security model proof has been started with partners of the project in [54] that allows to generate test needs associated to security properties verified on model.

6.4.2. Scenario-Based Verification and Validation

Participants: Fabrice Bouquet, Kalou Cabrera, Frédéric Dadeau, Elizabeta Fourneret.

Test scenarios represent an abstract test case specification that aims at guiding the model animation in order to produce relevant test cases. Contrary to the previous section, this technique is not fully automated since it requires the user to design the scenario, in addition to the model.

We have designed a scenario based testing language for UML/OCL that can be either connected to a model animation engine or to a symbolic animation engine, based on a set-theoretical constraint solver [22]. In the context of the ANR TASCCC project, we are investigating the automation of test generation from Security Functional Requirements (SFR), as defined in the Common Criteria terminology. SFRs represent security functions that have to be assessed during the validation phase of security products (in the project, the Global Platform, an operating system for latest-generation smart cards). To achieve that, we are working on the definition of description patterns for security properties, to which a given set of SFRs can be related. These properties are used to automatically generate test scenarios that produce model based test cases. The traceability, ensured all along the testing process, makes it possible to provide evidences of the coverage of the SFR by the tests, required by the Common Criteria to reach the highest Evaluation Assurance Levels.

We have proposed a dedicated formalism to express test properties. A test property is first translated into a finite state automaton which describes a monitor of its behaviors. We have proposed dedicated property coverage criteria that can be used either to measure the property coverage of a given test suite, or to generate test cases, exercising nominal or robustness aspects of the property.

In the context of the SecureChange project, we also investigate the evolution of test scenarios. As the system evolves, the model evolves, and the associated test scenarios may also evolve. We are currently extending the test generation and management of system evolutions to ensure the preservation of the security.

6.4.3. Mutation-based Testing of Security Protocols

Participants: Frédéric Dadeau, Pierre-Cyrille Héam.

Verification of security protocols models is an important issue. Nevertheless, the verification reasons on a model of the protocol, and does not consider its concrete implementation. While representing a safe model, the protocol may be incorrectly implemented, leading to security flaws when it is deployed. We have proposed a model-based penetration testing approach for security protocols [9]. This technique relies on the use of mutations of an original protocol, proved to be correct, for injecting realistic errors that may occur during the protocol implementation (e.g. re-use of existing keys, partial checking of received messages, incorrect formatting of sent messages, use of exponential/xor encryption, etc.). Mutations that lead to security flaws are used to build test cases, which are defined as a sequence of messages representing the behavior of the intruder. We have applied our technique on protocols designed in HLPSL, and implemented a protocol mutation tool that performs the mutations. The mutants are then analyzed by the CL-Atse [82] front-end of the AVISPA toolset [66]. Experiments show the relevance of the proposed mutation operators and the efficiency of the CL-Atse tool to conclude on the vulnerability of a protocol and produce an attack trace that can be used as a test case for implementations.

6.4.4. Code-related Test Generation and Static Analysis

Participants: Fabrice Bouquet, Frédéric Dadeau, Ivan Enderlin, Alain Giorgetti.

In collaboration with the CEA we enhance the innovative verification technique SANTE (Static ANalysis and TEsting), combining value analysis, program slicing and test generation, with two novel, optimized and adaptive strategies of program slicing based on threat dependencies [37]. We study the properties of threat dependencies, introduce the notion of slicing-induced cover, and prove the underlying theoretical results. Compared to a basic usage of program slicing, our advanced strategies need only quadratic additional work in order to optimize the calls of costly dynamic analysis. We give a detailed evaluation of all slicing strategies and compare them with one another.

We have designed a new annotation language for PHP, named PRASPEL for PHP Realistic Annotation SPEcification Language. This language relies on *realistic domains* which serve two purposes. First, they assign to a data a domain that is supposed to be specific w.r.t. a context in which it is employed. Second, they provide two features that are used for test generation: (i) *samplability* makes it possible to automatically generate a value that belongs to the realistic domain so as to generate test data, (ii) *predicability* makes it possible to check if the value belongs to a realistic domain. This approach is tool-supported in a dedicated framework for PHP which makes it possible to produce unit test cases using random data generators, execute the test cases on an instrumented implementation, and decide the conformance of the code w.r.t. the annotations by runtime assertion checking. This principle has been extended to generate grammar-based textual data [43] based on various strategies, namely uniform random generation, bounded exhaustive generation and rule-coverage-based test generation.

6.4.5. Specification, implementation and validation of generation algorithms

Participant: Alain Giorgetti.

We have shown how to use logic programming and bounded-exhaustive testing to design and validate algorithms generating a family of combinatorial objects [45]. The focus is on computer assistance for the task of validation of an implementation with respect to a different implementation or a formal specification. Among the numerous perspectives, these generation algorithms can to their turn be embedded in bounded exhaustive testing tools, such as the one proposed in [43].

6.5. Verification of Collaborative Systems

We investigate security problems occurring in decentralized systems. We develop general techniques to enforce read and update policies for controlling access to XML documents based on recursive DTDs (Document Type Definition). Moreover, we provide a necessary and sufficient condition for undoing safely replicated objects in order to enforce access control policies in an optimistic way.

6.5.1. Automatic Analysis of Web Services Security

Participants: Michaël Rusinowitch, Mathieu Turuani, Laurent Vigneron.

Automatic composition of web services is a challenging task. Many works have considered simplified automata models that abstract away from the structure of messages exchanged by the services. For the domain of secured services (using e.g. digital signing or timestamping) we propose a novel approach to automated orchestration of services under security constraints. Given a community of services and a goal service, we reduce the problem of generating a mediator between a client and a service community to a security problem where an intruder should intercept and redirect messages from the service community and a client service till reaching a satisfying state. In [30] we present a tool that compiles the attack trace desribing the execution of a the mediator into its corresponding runnable code. For that the tool computes an executable specification of the mediator as prudent as possible of her role in the orchestration. This specification is expressed in ASLan language, a formal language designed for modeling Web Services tied with security policies that was developed in AVANTSSAR project. Then we can check with automatic tools that this ASLan specification verifies required security properties such as secrecy and authentication. If no flaw is found, we compile the specification into a Java servlet that can be used by the mediator to execute the orchestration. This process has been implemented in AVANTSSAR Platform [29].

In [31] we give a decision procedure for the satisfiability problem of general deducibility constraints. Two cases are considered: the standard Dolev-Yao theory and its extension with an associative, commutative idempotent operator. The result is applied to solve the automated distributed orchestration problem for secured Web services.

Finall we show in [32] how to check satisfiability of negative deducibility constraints and we apply the result to the orchestration of secured services under non-disclosure policies. We show in particular how to handle separation-of-duty constraints in orchestration.

6.5.2. Secure Querying and Updating of XML Data

Participants: Abdessamad Imine, Houari Mahfoud, Michaël Rusinowitch.

Over the past years several works have proposed access control models for XML data where only read-access rights over nonrecursive DTDs are considered. A small number of works have studied the access rights for updates. In this work, we propose a general model for specifying access control on XML data in the presence of the update operations of W3C XQuery Update Facility [56], [48]. Our approach for enforcing such update specification is based on the notion of query rewriting. A major issue is that query rewriting for recursive DTDs is still an open problem [49], [55]. We show that this limitation can be avoided using only the expressive power of the standard XPath, and we propose a linear algorithm to rewrite each update operation defined over an arbitrary DTD (recursive or not) into a safe one in order to be evaluated only over the XML data which can be updated by the user. This work represents the first effort for securely XML updating in the presence of arbitrary DTDs (recursive or not) and a rich fragment of XPath. Finally, we study the interaction between read and update access rights to preserve the confidentiality and integrity of XML data.

We introduce an extension of hedge automata called bidimensional context-free hedge automata, proposing a new uniform representation of vertical and horizontal computation steps in unranked ordered trees. We also extend the parameterized rewriting rules used for modeling the W3C XQuery Update Facility in previous works, by the possibility to insert a new parent node above a given node. Since the rewrite closure of hedge automata languages with these extended rewriting systems is a computable context-free hedge language we can perform some static typechecking on these XML transformations [63].

6.5.3. On the Polling Problem in Social Networks

Participants: Bao Thien Hoang, Abdessamad Imine.

We tackle the polling problem in social networks where the privacy of exchanged information and user reputation are very critical. Indeed, users want to preserve the confidentiality of their votes and to hide, if any, their misbehaviors. Recent works proposed polling protocols based on simple secret sharing scheme and without requiring any central authority or cryptography system. But these protocols can be deployed safely provided that the social graph structure should be transformed into a ring-based structure and the number of participating users is perfect square. Accordingly, devising polling protocols regardless these constraints remains a challenging issue. In this work, we propose a simple decentralized polling protocol that relies on the current state of social graphs [58], [33]. More explicitly, we define one family of social graphs and show their structures constitute necessary and sufficient condition to ensure vote privacy and limit the impact of dishonest users on the accuracy of the output of the poll.

6.5.4. Access Control Models for Collaborative Applications

Participants: Fabrice Bouquet, Asma Cherif, Abdessamad Imine.

The importance of collaborative systems in real-world applications has grown significantly over the recent years. The most of new applications are designed in a distributed fashion to meet collaborative work requirements. Among these applications, we focus on Real-Time Collaborative Editors (RCE) that provide computer support for modifying simultaneously shared documents, such as articles, wiki pages and programming source code by dispersed users. Although such applications are more and more used into many fields, the lack of an adequate access control concept is still limiting their full potential. In fact, controlling access in a decentralized fashion in such systems is a challenging problem, as they need dynamic access changes and low latency access to shared documents. In [12], we propose a generic access control model based on replicating the shared document and its authorization policy at the local memory of each user. We consider the propagation of authorizations and their interactions. We propose an optimistic approach to enforce access control in existing collaborative editing solutions in the sense that the access control policy can be temporarily violated. To enforce the policy, we resort to the selective undo approach in order to eliminate the effect of illegal document updates. Since, the safe undo is an open issue in collaborative applications. We investigate a theoretical study of the undo problem and propose a generic solution for selectively undoing operations. Finally, we apply our framework on a collaboration prototype and measure its performance in the distributed grid GRID'5000 to highlight the scalability of our solution.

We realize the verification of Ramos protocol for concurrent writing and reconfiguration for collaborative systems in [23]. The Ramos protocol implements a fault-tolerant, and a context consistency (ensuring a total order of write operations) based on an asynchronous message-passing model. Communication takes place via gossip messages, which are sent at any frequency between a dynamic set of nodes into the collaborative system.

CELTIQUE Project-Team

5. New Results

5.1. Control-Flow Analysis by Abstract Interpretation

Control-flow analysis (CFA) of functional programs is concerned with determining how the program's functions call each other. In the case of the lambda calculus, this amounts to computing the flow of lambda expressions in order to determine what functions are effectively called in an application $(e_1\ e_2)$. This work shows that it is possible to use abstract interpretation techniques to derive systematically a control-flow analysis for a simple higher-order functional language. The analysis approximates the interprocedural control-flow of both function calls and returns in the presence of first-class functions and tail-call optimization. A number of advantages follow from taking this approach:

- The systematic derivation of a CFA for a higher-order functional language from a well-known operational semantics provides the resulting analysis with strong mathematical foundations. Its correctness follows directly from the general theorems of abstract interpretation.
- The approach is easily adapted to different variants of the source language. We demonstrate this by deriving a CFA for functional programs written in continuation-passing style.
- The common framework of these analyses enables their comparison. We take advantage of this to settle a question about the equivalence between the analysis of programs in direct and continuation-passing style.
- The resulting equations can be given an equivalent constraint-based presentation, providing *ipso facto* a rational reconstruction and a correctness proof of constraint-based CFA.

This work was published in the journal Information and Computation [14]

5.2. Secure the Clones: Static Enforcement of Policies for Secure Object Copying

Participants: Thomas Jensen, David Pichardie.

Exchanging mutable data objects with untrusted code is a delicate matter because of the risk of creating a data space that is accessible by an attacker. Consequently, secure programming guidelines for Java stress the importance of using defensive copying before accepting or handing out references to an internal mutable object.

However, implementation of a copy method (like clone()) is entirely left to the programmer. It may not provide a sufficiently deep copy of an object and is subject to overriding by a malicious sub-class. Currently no language-based mechanism supports secure object cloning.

We propose a type-based annotation system for defining modular copy policies for class-based object-oriented programs. A copy policy specifies the maximally allowed sharing between an object and its clone. We provide a static enforcement mechanism that will guarantee that all classes fulfill their copy policy, even in the presence of overriding of copy methods, and establish the semantic correctness of the overall approach in Coq.

The mechanism has been implemented and experimentally evaluated on clone methods from several Java libraries. The work as been presented at ESOP 2011. In 2012 a journal special issue has been published in Logical Methods in Computer Science [13].

5.3. A formally verified SSA-based middle-end

Participants: Delphine Demange, David Pichardie.

CompCert is a formally verified compiler that generates compact and efficient PowerPC, ARM and x86 code for a large and realistic subset of the C language. However, CompCert foregoes using Static Single Assignment (SSA), an intermediate representation that allows for writing simpler and faster optimizers, and is used by many compilers. In fact, it has remained an open problem to verify formally a SSA-based compiler middle-end.

We report on a formally verified, SSA-based, middle-end for CompCert. Our middle-end performs conversion from CompCert intermediate form to SSA form, optimization of SSA programs, including Global Value Numbering, and transforming out of SSA to intermediate form.

In addition to provide the first formally verified SSA-based middle-end, we address two problems raised by Leroy: giving a simple and intuitive formal semantics to SSA, and leveraging the global properties of SSA to reason locally about program optimizations. The work as been presented at ESOP 2012 [16].

5.4. Non linear analysis: fast inference of polynomial invariants

Participants: Thomas Jensen, David Cachera, Arnaud Jobin.

The problem of automatically inferring non-linear (polynomial) invariants of programs is still a challenge in program verification. A central observation in existing work on generating polynomial invariants is that n-ary relations between variables that can be described as the zeroes of a set of polynomials, correspond to a lattice of polynomials ideals. Such ideals are finitely generated , and all the approaches proposed so far in the literature rely on Gröbner base computations for computing ideal intersection or inclusion, or analysing the effects of polynomial assignments to variables. Computing Gröbner bases however slows down considerably the overall analysis.

We have proposed an abstract interpretation based method for inferring polynomial invariants that entirely avoids computing Gröbner bases. The method is precise and efficient, and is obtained without restricting the expressiveness of the polynomial programming language. Our analysis handles a general polynomial structured programming language that includes if and while constructs where branching conditions are both polynomial equalities and disequalities. Our analysis uses a form of weakest precondition calculus for showing that a polynomial relation $g = 0$ holds at the end of a program. We show that this backward approach, which was already observed to be well adapted to polynomial disequality guards can be extended to equality guards by using parameterized polynomial division.

Based on this anlysis, we have designed a constraint-based algorithm for inferring polynomial invariants. Such constraint-based techniques (rather than iteration) when dealing with loops means that it becomes feasible to analyse conditionals precisely, using parameterized polynomial division. A salient feature of this analysis, which distinguishes it from previous analyses, is that it does not require the use of Gröbner base computations. We have implemented this algorithm in Maple and our benchmarks show that our analyzer can successfully infer invariants on a sizeable set of examples, while performing two orders of magnitude faster than other existing implementations [19].

5.5. Result Certification of Static Analysis Results

Participants: Thomas Jensen, Frédéric Besson, Pierre-Emmanuel Cornilleau, Ronan Saillard.

Result Certification, Static program analysis, Decision procedures

We develop a lightweight approach for verifying *a posteriori* that the result of a static analysis is correct. The approach consists in encoding the program semantics directly inside an Intermediate Verification Language e.g., Why3 as an executable program interpreter. Running the standard VcGen of the IVL for the interpreter specialised for a program annotated with analysis results therefore amounts to generating program specific verification conditions [20]. This approach has the advantage of reducing the size of the Trusted Computing Base (TCB) because the VcGen is generic and language agnostic. Moreover, unlike traditional approaches, our TCB does not embed a compiler from the source code to the language of the IVL.

Verification conditions are usually discharged by Satisfiability Modulo Theory (SMT) provers that are therefore part of the TCB. To reduce further the TCB, we advocate for proof-generating SMT provers which results can be independently verified by reflexive Coq proof-checkers. For the EUF logic, we have proposed a novel compact format and proved correct an efficient Coq checker [17].

5.6. Towards efficient abstract domains for regular language based static analysis

Participants: Thomas Genet, Valérie Murat, Yann Salmon.

We develop a specific theory and the related tools for analyzing programs whose semantics is defined using term rewriting systems. The analysis principle is based on regular approximations of infinite sets of terms reachable by rewriting. The tools we develop use, so-called, Tree Automata Completion to compute a tree automaton recognizing a superset of all reachable terms. This over-approximation is then used to prove properties on the program by showing that some "bad" terms, encoding dangerous or problematic configurations, are not in the superset and thus not reachable. With such technique, like with any approximated technique, is when the "bad" terms are in the superset. We proposed a new CounterExample Guided Abstraction Refinement (CEGAR) algorithm for tree automata completion. Our approach relies on a new equational-abstraction based completion algorithm to compute a regular overapproximation of the set of reachable states in finite time. This set is represented by, so-called, R/E-automata, a new extended tree automaton formalism whose structure can be exploited to detect and remove false positives in an efficient manner. Our approach has been implemented in Timbuk and used to analyze Java programs by exploiting a translation from the Java byte code to term rewriting systems. These results have been published in [18]. Now, we aim at applying this technique to the static analysis of programming languages whose semantics is based on terms, like functional programming languages. The first step in this direction is to take into account the evaluation strategy of the language when approximating the set of reachable terms [30].

5.7. Cryptography

Participants: Pierre-Alain Fouque, Jean-Christophe Zapalowicz.

Pierre-Alain Fouque joined the team Celtique from September 2011 to August 2012. As a cryptographer, he still worked on symmetric cryptography with his PhD and postdoc students and proposed new security analysis of the block-ciphers AES and Camellia using meet-in-the-middle techniques in [27], [22] at IWSEC'12 and Indocrypt'12 and new security proofs for signature schemes AbdallaFLT12 at Eurocrypt'12 and elliptic-curve hash function [25] at LatinCrypt'12 with nice properties.

With Pierre-Alain, we also worked on more practical security aspects since his delegation in the Celtique team was to study side-channel attacks and formal methods. In side-channel attacks, we work with people from DGA and NTT in Japan to present new efficient attacks on one well-known implementation of RSA in many smartcards. Our attack targets any implementation of RSA using the Chinese Remainder Theorem in order to speed-up the computation, any exponentiation algorithm and the Montgomery multiplication. Usually, public-key cryptography requires large integer arithmetic and in order to accelerate the computation of the modulo, Montgomery proposed a new algorithm that avoids the need of arbitrary euclidean division which is the most consuming part of the exponentiation algorithm. This algorithm uses a small register (8, 16 or 32 bits depending on the architecture) during the computation and if a fault makes the value of this register much shorter, we show that we can recover the factorization of the RSA modulus in polynomial time. Furthermore, we describe on many proposed hardware architectures that our attack can indeed be used in practice if a laser is used to provoke the fault. This article has been published at CHES'12.

With people from DGA, we also studied how fault attack can be used to have buffer overflow effects. Indeed, by accelerating the clock, it is possible to avoid some instruction in the assembler code of a function. Consequently, if a fault avoids the function epilogue that restores the stack and registers to the state they were in before the function was called, then the stack pointer is changed and we can execute another function. Such attacks show that code executed in embedded processor have to be protected using buffer overflow techniques.

Finally, we also worked with people from DGA and Grenoble University to study security proofs in a computational logic. We show that the mode of operations of some hash functions is secure in [21] and published at CSF'12. In particular, we show a small bug in the security proof of the sponge construction used in the new SHA-3 candidate and winner of the competition Keccak.

COMETE Project-Team

6. New Results

6.1. Foundations of information hiding

Information hiding refers to the problem of protecting private information while performing certain tasks or interactions, and trying to avoid that an adversary can infer such information. This is one of the main areas of research in Comète; we are exploring several topics, described below.

6.1.1. Measuring information leakage

A fundamental concern in computer security is to control information flow, whether to protect confidential information from being leaked, or to protect trusted information from being tainted. In view of the pragmatic difficulty of preventing undesirable flows completely, there is now much interest in theories that allow information flow to be quantified, so that "small" leaks can be tolerated. In [19] we introduced g-leakage, a rich generalization of the min-entropy model of quantitative information flow. In g-leakage, the benefit that an adversary derives from a certain guess about a secret is specified using a gain function g. Gain functions allow a wide variety of operational scenarios to be modeled, including those where the adversary benefits from guessing a value close to the secret, guessing a part of the secret, guessing a property of the secret, or guessing the secret within some number of tries. We proved important properties of g-leakage, including bounds between min-capacity, g-capacity, and Shannon capacity. We also showed a deep connection between a strong leakage ordering on two channels, C1 and C2, and the possibility of factoring C1 into C2 C3 , for some C3 . Based on this connection, we proposed a generalization of the Lattice of Information from deterministic to probabilistic channels.

6.1.2. Interactive systems

In [12] we have considered systems where secrets and observables can alternate during the computation. We have shown that the information-theoretic approach which interprets such systems as (simple) noisy channels is not valid anymore. However, the principle can be recovered if we consider more complicated types of channels, that in Information Theory are known as channels with memory and feedback. We have shown that there is a complete correspondence between interactive systems and such kind of channels. Furthermore, we have shown that the capacity of the channels associated to such systems is a continuous function of the Kantorovich metric.

6.1.3. Unlinkability

Unlinkability is a privacy property of crucial importance for several systems (such as RFID or voting systems). Informally, unlinkability states that, given two events/items in a system, an attacker is not able to infer whether they are related to each other. However, in the literature we find several definitions for this notion, which are apparently unrelated and shows a potentially problematic lack of agreement. In [22] we shed new light on unlinkability by comparing different ways of defining it and showing that in many practical situations the various definitions coincide. It does so by (a) expressing in a unifying framework four definitions of unlinkability from the literature (b) demonstrating how these definitions are different yet related to each other and to their dual notion of "inseparability" and (c) by identifying conditions under which all these definitions become equivalent. We argued that the conditions are reasonable to expect in identification systems, and we prove that they hold for a generic class of protocols.

6.1.4. A compositional method to compute the sensitivity of differentially private queries

Differential privacy is a modern approach in privacy-preserving data analysis to control the amount of information that can be inferred about an individual by querying a database. The most common techniques are based on the introduction of probabilistic noise, often defined as a Laplacian parametric on the sensitivity of the query. In order to maximize the utility of the query, it is crucial to estimate the sensitivity as precisely as possible.

In [28] we considered relational algebra, the classical language for expressing queries in relational databases, and we proposed a method for computing a bound on the sensitivity of queries in an intuitive and compositional way. We used constraint-based techniques to accumulate the information on the possible values for attributes provided by the various components of the query, thus making it possible to compute tight bounds on the sensitivity.

6.1.5. A differentially private mechanism of optimal utility for a region of priors

Differential privacy (already introduced in the previous section) is usually achieved by using mechanisms that add random noise to the query answer. Thus, privacy is obtained at the cost of reducing the accuracy, and therefore the utility, of the answer. Since the utility depends on the user's side information, commonly modeled as a prior distribution, a natural goal is to design mechanisms that are optimal for every prior. However, it has been shown in the literature that such mechanisms do not exist for any query other than counting queries.

Given the above negative result, in [38] we considered the problem of identifying a restricted class of priors for which an optimal mechanism does exist. Given an arbitrary query and a privacy parameter, we geometrically characterized a special region of priors as a convex polytope in the priors space. We then derived upper bounds for utility as well as for min-entropy leakage for the priors in this region. Finally we defined what we call the tight-constraints mechanism and we discussed the conditions for its existence. This mechanism has the property of reaching the bounds for all the priors of the region, and thus it is optimal on the whole region.

6.1.6. Differential privacy with general metrics

Differential privacy, already described above, is a formal privacy guarantee that ensures that sensitive information relative to individuals cannot be easily inferred by disclosing answers to aggregate queries. If two databases are adjacent, i.e. differ only for an individual, then querying them should not allow to tell them apart by more than a certain factor. The transitive application of this property induces a bound also on the distinguishability of two generic databases, which is determined by their distance on the Hamming graph of the adjacency relation.

In [37] we lifted the restriction relative to the Hamming graphs and we explored the implications of differential privacy when the indistinguishability requirement depends on an arbitrary notion of distance. We showed that we can express, in this way, (protection against) kinds of privacy threats that cannot be naturally represented with the standard notion. We gave an intuitive characterization of these threats in terms of Bayesian adversaries, which generalizes the characterization of (standard) differential privacy from the literature. Next, we revisited the well-known result on the non-existence of universally optimal mechanisms for any query other than counting queries. We showed that in our setting, for certain kinds of distances, there are many more queries for which universally optimal mechanisms exist: Notably sum, average, and percentile queries. Finally, we showed some applications in various domains: statistical databases where the units of protection are groups (rather than individuals), geolocation, and smart metering.

6.1.7. Privacy for location-based systems

The growing popularity of location-based systems, allowing unknown/untrusted servers to easily collect and process huge amounts of users' information regarding their location, has recently started raising serious concerns about the privacy of this kind of sensitive information. In [36] we studied geo-indistinguishability, a formal notion of privacy for location-based systems that protects the exact location of a user, while still allowing approximate information - typically needed to obtain a certain desired service - to be released.

Our privacy definition formalizes the intuitive notion of protecting the user's location within a radius r with a level of privacy that depends on r. We presented three equivalent characterizations of this notion, one of which corresponds to a generalized version [37] of the well-known concept of differential privacy. Furthermore, we presented a perturbation technique for achieving geo-indistinguishability by adding controlled random noise to the user's location, drawn from a planar Laplace distribution. We demonstrated the applicability of our technique through two case studies: First, we showed how to enhance applications for location-based services with privacy guarantees by implementing our technique on the client side of the application. Second, we showed how to apply our technique to sanitize location-based sensible information collected by the US Census Bureau.

6.1.8. Compositional analysis of information hiding

Systems concerned with information hiding often use randomization to obfuscate the link between the observables and the information to be protected. The degree of protection provided by a system can be expressed in terms of the probability of error associated to the inference of the secret information. In [15] we considered a probabilistic process calculus to specify such systems, and we studied how the operators affect the probability of error. In particular, we characterized constructs that have the property of not decreasing the degree of protection, and that can therefore be considered safe in the modular construction of these systems. As a case study, we applied these techniques to the Dining Cryptographers, and we derived a generalization of Chaum's strong anonymity result.

In [29], a similar framework was proposed for reasoning about the degree of differential privacy provided by such systems. In particular, we investigated the preservation of the degree of privacy under composition via the various operators. We illustrated our idea by proving an anonymity-preservation property for a variant of the Crowds protocol for which the standard analyses from the literature are inapplicable. Finally, we made some preliminary steps towards automatically computing the degree of privacy of a system in a compositional way.

6.1.9. Anonymous and route-secure communication systems

Incentives to Cooperation. Anonymity systems have a broad range of users, ranging from ordinary citizens who want to avoid being profiled for targeted advertisements, to companies trying to hide information from their competitors, to entities requiring untraceable communication over the Internet. With these many potential users, it would seem that anonymity services based on a consumer/provider users will naturally be well-resourced and able to operate efficiently. However, cooperation cannot be taken for granted. Current deployed systems show that some users will indeed act selfishly, and only use the system to send their messages whilst ignoring the requests to forward others' messages. Obviously, with not enough cooperative users, the systems will hardly operate at all, and will certainly not be able to afford adequate anonymity guarantees. It is therefore vital that these systems are able to deploy incentives to encourage users' cooperation and so make the anonymity provision effective. Some interesting approaches to achieve that have been proposed, such as make running relays easier and provide better forwarding performance.

To evaluate whether these approaches are effective, we need a framework which empowers us to analyze them, as well as provide guidelines and some mechanism design principles for incentive schemes. This much we have provided in [30], exploiting notions and techniques from Game Theory. We proposed a game theoretic framework and used it to analyze users' behaviours and also predict what strategies users will choose under different circumstances and according to their exact balance of preferences among factors such as anonymity, performance (message delivery time) and cost. Significantly, we also used the model to assess the effectiveness of the gold-star incentive mechanism, which was introduced in Tor network to encourage users to act as cooperative relays, and thus enhance the service performance for well-behaved forwarders.

Trust in anonymity networks. Trust metrics are used in anonymity networks to support and enhance reliability in the absence of verifiable identities, and a variety of security attacks currently focus on degrading a user's trustworthiness in the eyes of the other users. In [16] we have presented an enhancement of the Crowds anonymity protocol via a notion of trust which allows crowd members to route their traffic according to their perceived degree of trustworthiness of each other member of the crowd. Such trust relations express a measure of an individual's belief that another user may become compromised by an attacker, either by a direct attempt to corrupt or by a denial-of-service attack. Our protocol variation has the potential of improving the overall trustworthiness of data exchanges in anonymity networks, which cannot normally be taken for granted in a context where users are actively trying to conceal their identities. Using such formalization, in the paper we have then analyzed quantitatively the privacy properties of the protocol under standard and adaptive attacks.

6.2. Foundations of Concurrency

Distributed systems have changed substantially in the recent past with the advent of phenomena like social networks and cloud computing. In the previous incarnation of distributed computing the emphasis was on consistency, fault tolerance, resource management and related topics; these were all characterized by *interaction between processes*. Research proceeded along two lines: the algorithmic side which dominated the Principles Of Distributed Computing conferences and the more process algebraic approach epitomized by CONCUR where the emphasis was on developing compositional reasoning principles. What marks the new era of distributed systems is an emphasis on managing access to information to a much greater degree than before.

The *Concurrent constraint programming (ccp)* paradigm focuses on information access and therefore it is suited for this new era of concurrent systems. Ccp singles out the fundamental aspects of asynchronous systems whose agents (or processes) evolve by accessing information in a global medium. In the works [20], [21], [31], [26] described below we developed algorithms and extended the foundations of ccp.

6.2.1. Spatial and Epistemic Modalities for Constraint-based Calculi

Epistemic concepts were crucial in distributed computing as was realized in the mid 1980s with Halpern and Moses' groundbreaking paper on common knowledge. This led to a flurry of activity in the next few years with many distributed protocols being understood from an epistemic point of view. The impact of epistemic ideas in the concurrency theory community was slower in coming. We believe that epistemic ideas need to be exploited more by concurrency theorists and we did so in the following works.

In [26] we introduced spatial and epistemic process calculi for reasoning about spatial information and knowledge distributed among the agents of a system. We also introduced domain-theoretical structures to represent spatial and epistemic information. Finally we provided operational and denotational techniques for reasoning about the potentially infinite behaviour of spatial and epistemic processes. We also gave compact representations of infinite objects that can be used by processes to simulate announcements of common knowledge and global information. We also developed an interpreter of these calculi in [31].

6.2.2. Bisimilarity for Constraint-based Calculi

Bisimilarity is a standard behavioural equivalence in concurrency theory, but a well-behaved notion of bisimilarity for ccp has been proposed only recently. When the state space of a system is finite, the ordinary notion of bisimilarity can be computed via the well-known partition refinement algorithm, but unfortunately, this algorithm does not work for ccp bisimilarity. In [20] we proposed a variation of the partition refinement algorithm for verifying ccp bisimilarity. To the best of our knowledge this is the first work providing for the automatic verification of program equivalence for ccp.

In [20] we only studied the strong version of bisimilarity. Weak bisimiliarity is obtained from the strong case by taking into account only the actions that are observable in the system. Typically, the standard partition refinement can also be used for deciding weak bisimilarity simply by using Milner's reduction from weak to strong bisimilarity; a technique referred to as saturation. In [21] we showed that, because of its involved labeled transitions, the above-mentioned saturation technique does not work for ccp. We also gave an alternative reduction from weak ccp bisimilarity to the strong one that allows us to use the ccp partition refinement algorithm for deciding this equivalence.

In the more traditional setting of the pi-calculus we have also proposed an approach to restrict access to information.

6.2.3. Locality in the Pi-Calculus

In [25] we enriched the pi-calculus with an operator for confidentiality (hide), whose main effect is to restrict the access to the object of the communication, thus representing confidentiality in a natural way. The hide operator is meant for local communication, and it differs from new in that it forbids the extrusion of the name and hence has a static scope. Consequently, a communication channel in the scope of a hide can be implemented as a dedicated channel, and it is more secure than one in the scope of a new. To emphasize the difference, we introduced a spy context that represents a side-channel attack and breaks some of the standard security equations for new. To formally reason on the security guarantees provided by the hide construct, we also introduced an observational theory and establish stronger equivalences by relying on a proof technique based on bisimulation semantics.

6.2.4. Foundations of Probabilistic Concurrent Systems

In [17] we have solved an open problem in the literature by proving that two known semantics for the probabilistic mu-calculus, a denotational semantics and a two-player stochastic game semantics, coincide on all models.

In [18] we have improved the result of [17] by introducing a new logic called probabilistic mu-calculus with independent product. We have proved that two semantics coincide in all models: a denotational semantics and a two-player game semantics based on a novel class of concurrent games. Furthermore, we have shown how the new logic is strictly more expressive than the other. This allows the encoding of other important temporal logics for probabilistic concurrent systems such as PCTL.

In [27] we have introduced a proof system designed for supporting human-aided verification of properties (expressed as probabilistic mu-calculus formulas ([17])) of concurrent probabilistic processes described by SOS-style operational semantics.

6.2.5. Interference metrics for Mobile ad-hoc networks (MANETs)

Mobile ad-hoc networks consist of a collection of nodes that communicate with each other through wireless links without a pre-established networking infrastructure. A common feature of most of these networks is free node mobility. Each device will therefore change its links to other devices frequently. These frequent changes in the network topology can cause the nodes to continuously enter and exit each other transmission area. Hence, highly dynamic routing algorithms are needed to ensure the connectivity. Moreover, mobile devices may have strict requirements on the energy consumption because their expected life-time often depends on the energy stored in a battery or other exhaustible power sources. For these reasons, finding a good trade-off between network connectivity, power saving and interference reduction is one of the most critical challenges in managing mobile ad hoc networks. In [23], we have proposed an effective framework for analysing protocol connectivity and measuring the level of interference and, based on that for developing novel interference-aware communication strategies. Though other models exist in the literature, to our best knowledge, our framework is the most comprehensive and effective for the behavioral analysis and a quantitative assessment of interference for wireless networks in the presence of node mobility.

COMPSYS Project-Team

6. New Results

6.1. Enhancing the Compilation of Synchronous Data-Flow Languages

Participants: Paul Feautrier, Abdoulaye Gamatié [LIFL], Laure Gonnord [Compsys/LIFL].

In [25] a new (light) numerical-Boolean abstraction was proposed for an efficient static analysis of synchronous programs that describe multi-clock embedded systems in the language Signal. In this abstraction, relations between clocks and numerical variables are modeled by Boolean-affine formulas. These formulas can easily be extracted from the program text. From the results of a satisfiability test of these formulas, clock properties can be deduced, which, when submitted to the Signal compiler, may improve the resulting target program.

In collaboration with Abdoulaye Gamatié, we proposed an extension of the previous approach to modular programs. This extension necessitates the use of an extended SMT (satisfiability modulo theory) solver – able for instance to deal with quantifier elimination – which has been implemented by Paul Feautrier by reusing some of the Syntol tools. This work is still unpublished but will be soon submitted to a journal.

6.2. Dataflow Analysis of Polyhedral X10 Programs

Participants: Paul Feautrier, Sanjay Rajopadhye [Colorado State Universty], Vijay Saraswat [IBM Research], Tomofumi Yuki [Colorado State University].

X10 is a recent parallel language, developed by IBM Research, whose aim is to increase programmers productivity. It is a descendant of Java and it includes several new parallel constructs, such as `async` and `finish`, which generalize `fork` and `join`, clocks, which generalize barriers, and `at`, which enables the remote execution of program fragments. X10 programs are guaranteed to be deadlock-free, but may exhibit non-deterministic behaviors or *races*.

We have devised a verifier for the `async/finish` fragment of X10 and polyhedral programs. The approach consists in computing the source of each array access. In the presence of parallel constructs, the sequencing predicate is no longer a total order and a read may have several sources, which indicates a race. A proof-of-concept tool has been implemented. This work will be presented at the next *Principles and Practice of Parallel Programming* conference (PPoPP'13 in Shenzen, China) [13].

6.3. Data-Aware Process Networks

Participants: Christophe Alias, Alexandru Plesco [Compsys/Zettice].

New techniques were introduced to generate and compile optimized data-aware process networks, from a C program annotated with pragmas (see Section 5.9 and the software tool Dcc). These techniques are essential for the Zettice start-up and are not made publicly available for the moment.

6.4. Optimizing Remote Accesses for HLS

Participants: Christophe Alias, Alain Darte, Alexandru Plesco [Compsys/Zettice].

Some data- and compute-intensive applications can be accelerated by offloading portions of codes to platforms such as GPGPUs or FPGAs. However, to get high performance for these kernels, it is mandatory to restructure the application, to generate adequate communication mechanisms for the transfer of remote data, and to make good usage of the memory bandwidth. In the context of the high-level synthesis (HLS), from a C program, of hardware accelerators on FPGA, we showed how to automatically generate optimized remote accesses for an accelerator communicating to an external DDR memory. Loop tiling is used to enable block communications, suitable for DDR memories. Pipelined communication processes are generated to overlap communications and computations, thereby hiding some latencies, in a way similar to double buffering. Finally, not only intra-tile but also inter-tile data reuse is exploited to avoid remote accesses when data are already available in the local memory.

We showed how to generate the sets of data to be read from (resp. written to) the external memory just before (resp. after) each tile so as to reduce communications and reuse data as much as possible in the accelerator. The main difficulty arises when some data may be (re)defined in the accelerator and should be kept locally. We proposed an automatic optimized code generation scheme, entirely at source-level, i.e., in C, that allows us to compile all the necessary glue (the communication processes) with the same HLS tool as for the computation kernel. Our method, implemented in the tool Chuba (see Section 5.7) uses advanced polyhedral techniques for program analysis and transformation. Experiments with Altera HLS tools demonstrate how to use our techniques to efficiently map C kernels to FPGA.

This work, astride two different fields (compilation for high-performance computing and high-level synthesis) turned out to be very difficult to publish. It was finally accepted at PPoPP'12 [6], but only as a short paper (2 pages). We requested to retain the copyright of this work to be able to publish a longer version. It was accepted at the IMPACT'12 workshop [7], which makes paper available on the web, but with no copyright. It was finally accepted as a full publication at the DATE'13 conference [8].

6.5. Parametric Inter-Tile Reuse for Kernel Offloading

Participants: Alain Darte, Alexandre Isoard.

The method described in Section 6.4 is not parametric in terms of the tile size, i.e., the tile size needs to be fixed before compiling the program. Furthermore, the size of the required local memory depends on the tile size and is available only after program analysis. As a result, to select the tile size with respect to the size of the local memory, the program first needs to be compiled (actually analyzed) for all tile sizes. A parametric program analysis would be much more convenient. The situation is even worse to get the runtime performances in terms of the tile size. Indeed, so far, Chuba generates a C code that is generic and cannot be immediately compiled by C2H. A few modifications by hand are still needed, such as inserting the adequate pragmas for C2H, transforming array accesses to linearized addresses with the right base addresses, changing some arrays into non-aliasing pointers so that C2H, whose dependence analyzer and software pipeliner are weak, can generate codes with the right initiation intervals, etc. These changes are minor and systematic and take time when performed by hand. A fully-parametric compilation scheme would be a plus.

The formulation proposed in Section 6.4 is unfortunately quadratic in terms of the tile size, which prevents to parameterize it. Indeed, it relies on parametric linear programming, which works only with a linear use of parameters. As part of the Master internship of Alexandre Isoard, we nevertheless succeeded to design a fully-parametric scheme for inter-tile reuse and buffer size computation. The method is much more involved but is still compatible with approximations. These results have still to be implemented and submitted for publication.

6.6. Semantic Program Transformations

Participants: Christophe Alias, Guillaume Iooss, Sanjay Rajopadhye [Colorado State University].

Traditionally, a program transformation is considered to be *correct* if each data dependence of the original program is respected. In that case, both original and transformed programs perform *exactly* the same computation. We can relax this condition by expecting both programs to perform the same computation, modulo the *semantic* properties of the operators (e.g., associativity, commutativity). Semantic program transformations extend the traditional corpus of program transformations and can reveal new optimization opportunities.

More specifically, we are interested in *semantic loop tiling*, a special case of loop tiling, where the input arrays are tiled, and the program is restructured to use high-level matrix operations between data tiles, instead of the original scalar operations. Surprisingly, it turns out that in most cases, the semantic tiling is simply obtained by substituting the scalar variables by the tiles (matrices), and the original operators by the corresponding matrix operators (e.g., a/b by MatMul(A,Inv(B))). The approach currently investigated consists in two steps: (i) guess the semantic tiling, and (ii) prove the (semantic) equivalence with the original program.

Our current contribution is an heuristic to check the equivalence of two programs modulo associativity/commutativity so as to achieve the step (ii). The two programs should fit in the polyhedral model but can involve explicit reductions. This work is currently under submission, and is part of the PhD thesis of Guillaume Iooss.

6.7. Modular Termination of Large Programs

Participants: Christophe Alias, Guillaume Andrieu [LIFL], Laure Gonnord [Compsys/LIFL].

Program termination is an essential step in program verification. In [16], we showed how to check the termination of programs whose control can be summarized by an integer interpreted automaton. This was done by computing a *ranking function* (kind of schedule) by means of integer linear programming techniques. This approach, though powerful, clearly lacks scalability and cannot handle large programs.

We overcame this limitation by proceeding into two steps. First, we extract, from the program to be analyzed, the part useful for termination, i.e., the smaller program slice with the same control behavior. Then, we show that proving the termination of the whole program (slice) boils down to prove the termination of small programs, which can be handled by the technique of [16]. Experimental results show that many large programs can be handled this way.

This work was part of the engineer internship of Guillaume Andrieu. Our technique has been implemented in a tool called SToP (see Section 5.12) and presented at the workshop TAPAS'12 [9].

6.8. Lower Bounds for the Inherent Data Locality Properties of Computations

Participants: Venmugil Elango [OSU, Columbus, USA], Louis-Noël Pouchet [UCLA, Los Angeles, USA], P. Sadayappan [OSU, Columbus, USA], J. (Ram) Ramanujam [LSU, Houston, USA], Fabrice Rastello.

Data movement will account for most of the energy as well as execution time on upcoming exascale architectures, including data movement between processors as well as data movement across the memory hierarchy within each processor. Therefore a fundamental characterization of the data access complexity of algorithms is increasingly important.

We addressed the problem of data access or I/O complexity in a two-level memory hierarchy, as studied in the seminal work of Hong and Kung [26]. We developed a novel approach based on graph min-cut for deriving lower bounds on I/O complexity with two significant advantages over the S-partitioning model of Hong and Kung: (1) the approach can be used to develop analytical expressions with tighter lower bounds for I/O, and (2) unlike any previous model, our new lower bound approach can be automated for analyzing an arbitrary computational directed acyclic graph. We show tighter analytically-derived lower bounds as well as very promising experimental results thanks to a prototype tool that implements our fully-automated analysis.

This work has been submitted and is part of an informal collaboration with P. Sadayappan from the University of Columbus (CSU).

6.9. A Polynomial Spilling Heuristic: Layered Allocation

Participants: Albert Cohen [Inria, Parkas], Boubacar Diouf [Université Paris Sud, Parkas], Fabrice Rastello.

Register allocation is subdivided into two sub-problems: first, the *allocation* (or its dual problem the *spilling*) selects the set of variables that will reside in registers (resp. in memory) at each point of the program. Then, the *assignment* or *coloring* picks a specific register where a variable will reside. Building on some properties of the static single assignment form (SSA), it is now possible to decouple the allocation from the assignment. Indeed, the interference graph of a program in SSA form is a chordal graph. In this context, MAXLIVE, the maximal number of variables simultaneously live at a program point, is used during the spilling phase as a criterion to guarantee that the forthcoming assignment will be performed without any spill. If MAXLIVE is lower than or equal to R, the number of available registers, then all the variables will be assigned without any spill. This *decoupled* approach was advocated by Fabri, Appel and George, Darte et al., and others.

Existing spilling heuristics rely on a sufficient condition to guarantee register assignment, and incrementally spill until the condition holds. As we just mentioned, for programs under SSA, the condition is necessary and sufficient: MAXLIVE has to be lower than or equal to R. Incremental spilling decisions to satisfy this condition tend to be overly local and suboptimal. Indeed, incremental spilling itself is NP-complete, and heuristics based upon it trade too much their optimality for polynomiality. In contrast to incremental spilling, we proposed to adopt the symmetric approach: incremental allocation. Intuition for it emerges from two observations allowing for more global spilling decisions:

1. Register allocation is pseudo-polynomial in the number of registers, suggesting a heuristic that solves (optimally) roughly $R/step$ allocation problems on $step$ registers each. The final allocation is the layered composition of the stepwise allocations.
2. Stepwise optimality does not guarantee an overall optimal allocation, but experiments show that it comes very close to optimal, even with $step = 1$. Intuition for this comes from recent work by Diouf et al., observing that allocation decisions tend to be a monotonic function of the number of registers.

This work, which will be presented at CGO'13 [11], proposes a new graph-based allocation heuristic, based on a maximum clique cover formulation to define the profitability of spilling variables. It exploits the pseudo-polynomial complexity in the number of registers of the allocation problem under SSA — as opposed to the symmetric, spilling problem, which remains strongly NP-complete. It addresses the spill-everywhere problem in a decoupled context and also proposes an extension to non-decoupled approaches. It introduces *layered allocation* a new strategy that incrementally allocates variables instead of incrementally spilling variables. The evaluation performed on standard benchmarks shows that this new approach is near-optimal.

6.10. Interaction Between Spilling and Scheduling

Participants: Quentin Colombet, Alain Darte, Fabrice Rastello.

As explained in Section 6.9 , it is possible to decouple the register allocation problem in two successive phases: a first *spilling* phase places `load` and `store` instructions so that the register pressure at all program points is small enough, a second *assignment* and *coalescing* phase maps the remaining variables to physical registers and reduces the number of `move` instructions among registers. At CASES'11 [18], we presented a new integer linear programming (ILP) formulation, for load-store architectures, to capture "optimal" spilling in a more accurate and more expressive way than previous approaches. Among other features, we can express SSA ϕ-functions, memory-to-memory copies, and the fact that a value can be stored simultaneously in a register and in memory.

We used this ILP formulation to experimentally analyze the impact of the different heuristic strategies and compare them with optimal solutions. While "optimal" solutions show significant improvements for static spill costs, it turned out that runtime performances were disappointing (if not random). We conducted various experiments to understand this behavior and discovered that the interaction with scheduling is actually higher than expected. Micro-architectural features (e.g., memory latencies that can be hidden by prefetching, bundling that can hide cycles) have to be accounted for in the model, which is never done. These experiments and analysis are described in Chapter 4 of Quentin Colombet's PhD thesis [1].

6.11. Elimination of Parallel Copies Using Code Motion on Data Dependence Graphs

Participants: Florian Brandner, Quentin Colombet.

Traditional approaches to copy elimination during register allocation are based on interference graphs and register coalescing. Variables are represented as nodes in a graph, which are coalesced, if they can be assigned the same register. However, decoupled approaches strive to avoid interference graphs and thus often resort to local recoloring.

A common assumption of existing coalescing and recoloring approaches is that the original ordering of the instructions in the program is not changed. We developed an extension of a local recoloring technique called Parallel Copy Motion. We perform code motion on data dependence graphs in order to eliminate useless copies and reorder instructions, while at the same time a valid register assignment is preserved. Our results show that even after traditional register allocation with coalescing our technique is able to eliminate an additional 3% (up to 9%) of the remaining copies and reduce the weighted costs of register copies by up to 25% for the SPECINT 2000 benchmarks. In comparison to Parallel Copy Motion, our technique removes 11% (up to 20%) more copies and up to 39% more of the copy costs.

These results have been accepted for publication at SAC'12 [10] and, in a longer version, in the journal Computer Languages, Systems, and Structures [5].

CONTRAINTES Project-Team

6. New Results

6.1. Inferring Reaction Rule Models from Ordinary Differential Equations

Participants: François Fages, Steven Gay, Sylvain Soliman.

Many models in Systems Biology are described as Ordinary Differential Equations (ODEs), which allow for numerical integration, bifurcation analyses, parameter sensitivity analyses, etc. However, before fixing the kinetics and parameter values and going to simulations, various analyses can be performed based only on the structure of the model. This approach has rapidly developed in Systems Biology in the last decade, with for instance, the analyses of structural invariants in Petri net representation, model reductions by subgraph epimorphims, qualitative attractors in logical dynamics or temporal logic properties by analogy to circuit and program verification. These complementary analysis tools do not rely on kinetic information, but on the structure of the model with reactions.

In [8], [19], we present a symbolic computation algorithm for inferring a reaction model from an ODE system, based a general compatibility condition between the kinetic expression and the structure of a reaction, and report on its use for automatically curating the writing in SBML of the models in the respository biomodels.net. SBML is now a standard for sharing and publishing reaction models. However, since SBML does not enforce any coherence between the structure and the kinetics of a reaction, an ODE model can be transcribed in SBML without reflecting the real structure of the reactions, hereby invalidating many structural analyses. We show that the automatic writing in SBML of the models of biomodels.net allows us to reduce the percentage of models with a non well-formed reaction from 66% to 28%.

6.2. Petri Net Analyses of Biochemical Networks using Constraint Logic Programming

Participants: François Fages, Thierry Martinez, Faten Nabli, Sylvain Soliman.

Petri nets are a simple formalism for modeling concurrent computation. Recently, they have emerged as a promising tool for modeling and analyzing biochemical interaction networks, bridging the gap between purely qualitative and quantitative models. Biological networks can indeed be large and complex, which makes their study difficult and computationally challenging.

In [10], we focus on two structural properties of Petri nets, siphons and traps, that bring us information about the persistence of some molecular species. We present a Boolean model and two constraint-based methods for enumerating all minimal siphons and traps of a Petri net, by iterating the resolution of Boolean satisfiability problems executed with either a SAT solver or a CLP(B) program. We compare the performances of these methods with respect to a state-of-the-art algorithm from the Petri net community. On a benchmark with 80 Petri nets from the Petriweb database and 403 Petri nets from curated biological models of the Biomodels database, we show that miniSAT and CLP(B) solvers are overall both faster by two orders of magnitude with respect to the dedicated algorithm. Furthermore, we analyse why these programs perform so well on even very large biological models and show a polynomial time complexity result for Petri nets of fixed treewidth, using a similar theorem for constraint satisfaction problems with bounded treewidth constraint graphs.

In [5] we present a method to compute the minimal semi-positive invariants of a Petri net representing a biological reaction system, as resolution of a Constraint Satisfaction Problem. This analysis brings both qualitative and quantitative information on the models, in the form of conservation laws, consistency checking, etc. thanks to finite domain constraint programming. It is noticeable that some of the most recent optimizations of standard invariant computation techniques in Petri nets correspond to well-known techniques in constraint solving, like symmetry-breaking. A simple implementation based on GNU-Prolog's finite domain solver, and including symmetry detection and breaking, was incorporated into the BIOCHAM modelling environment and in the independent tool Nicotine. Some illustrative examples and benchmarks are provided.

6.3. Subgraph Epimorphisms

Participants: François Fages, Steven Gay, Thierry Martinez, Francesco Santini, Sylvain Soliman.

The operations of deleting and merging vertices are natural operations for reducing a graph. While graph reductions through a sequence of vertex deletions (resp. mergings) characterize subgraph isomorphisms (resp. graph epimorphisms), sequences of both vertex deletion and merging operations characterize subgraph epimorphisms. Our proposal is thus to use subgraph epimorphism for comparing graphs in applications in systems biology and image analysis, when a more flexible notion than the classical notion of subgraph isomorphism is required.

In collaboration with Christine Solnon (INSA Lyon), we have developed the theory of subgraph epimorphisms. We have defined the SEPI, EPI and SISO distances between two graphs as the size of the largest SEPI (resp. EPI, SISO) lower bound graphs. These distances are equal to the minimum number of respectively vertex deletion and/or merging operations that are necessary to obtain isomorphic graphs. They are also metrics on graphs and we have $d_d \geq d_{md}$ and $d_m \geq d_{md}$. From a computational point of view, we have shown that the existence of a SEPI between two graphs is an NP-complete problem and have presented a constraint satisfaction algorithm for solving it.

Our algorithm is implemented in BIOCHAM and is currently improved for better performance on large graphs and generalized as a SEPI graph constraint propagation algorithm for computing SEPI lower and upper bounds.

6.4. Parameter Search with Temporal Logic Constraints

Participants: Grégory Batt, François Fages, Anthony Lins, Sylvain Soliman, Pauline Traynard, Jannis Uhlendorf, Luma Vittorino.

Our method for solving temporal logic constraints in first-order linear time logic LTL(R_{lin}), opens up the field of model-checking to optimization through the definition of a continuous degree of satisfaction for temporal logic formulae. This satisfaction degree can be used in a number of ways, e.g. as a fitness function with continuous optimization methods to find unknown parameter values in a model, to perform sensitivity analyses and compute the robustness of a system w.r.t. a temporal property and a perturbation of the parameters. or to find control parameters.

This approach is implemented in BIOCHAM and is one unique feature of this modeling environment. In this implementation, the continuous optimization procedure we use is the Covariance Matrix Adaptation Evolutionary Strategy CMAES of Nikolaus Hansen from the EPI TAO. A parallel version of Biocham implements this method on the Jade cluster of 10000 cores at GENCI for running our most challenging parameter search problems.

This year, in collaboration with Fernando Buarque, we have explored another continuous optimization method of the family of Particle Swarm Optimization (PSO), called Fish School Optimization (FSS). In [13], we report on our first results which are encouraging for using FSS for decreasing the sensitivity of the method to initial conditions and being able to maintain several swarms of solutions.

6.5. Coupled Model of the Cell Cycle and Circadian Clock

Participants: François Fages, Sylvain Soliman, Denis Thieffry, Pauline Traynard.

Recent advances in cancer chronotherapy techniques support the evidence that there exist imortant links between the cell cycle and the circadian clock genes. One purpose for modeling these links is to better understand how to efficiently target malignant cells depending on the phase of the day and patient characterictics. This is at the heart of our participation in collaboration with the EPI BANG in the EraNet SysBio project C5Sys, follow up of the former EU STREP project TEMPO.

This year we have investigated the effect of transcription inhibition during mitosis, as a reverse coupling from the cell cycle to the circadian clock. We use temporal logic constraints and the parallel version of BIOCHAM for parameter search, running on the Jade cluster of 10000 processors at the GENCI CINES, to couple dynamical models in high dimension and fit models to experimental data time series obtained in Franck Delaunay's lab in Nice, CNRS.

6.6. STL-based Analysis of TRAIL-induced Apoptosis

Participants: Grégory Batt, François Bertaux, Szymon Stoma.

Extrinsic apoptosis is a programmed cell death triggered by external ligands, such as the TNF-related apoptosis inducing ligand (TRAIL). Depending on the cell line, the specific molecular mechanisms leading to cell death may significantly differ. Precise characterization of these differences is crucial for understanding and exploiting extrinsic apoptosis. Cells show distinct behaviors on several aspects of apoptosis, including (i) the relative order of caspases activation, (ii) the necessity of Mitochondria Outer Membrane Permeabilization (MOMP) for effector caspase activation, and (iii) the survival of cell lines overexpressing Bcl2, leading to classification of cell lines into two groups (type I and type II). In [21], we challenge this type I/II cell line classification. We encode the three aforementioned distinguishing behaviors in a formal language, called signal temporal logic (STL), and use it to extensively test the validity of a previously-proposed model of TRAIL-induced apoptosis with respect to experimental observations made on different cell lines. Then, STL-guided parameter search is used to solve the few inconsistencies found between model and data. We show that these three criteria do not define consistent cell line classifications in type I or type II, and suggest mutants that are predicted to exhibit ambivalent behaviors. In particular, this finding sheds light on the role of a feedback loop between caspases, and reconciliates two apparently-conflicting views regarding the importance of either upstream or downstream processes for cell type determination. More generally, our work suggests that rather than being considered as defining criteria for cell type classification, these three distinguishing behaviors should be merely considered as type I or II features. On the methodological point of view, this work illustrates the biological relevance of STL-diagrams, STL population data, and STL-guided parameter search. Such tools are well adapted to the ever-increasing availability of heterogeneous knowledge on complex signal transduction pathways.

6.7. Real-time Control of Gene Expression in Yeast

Participants: Grégory Batt, François Fages, Jannis Uhlendorf, Jean-Baptiste Lugagne, Artémis Llamosi, Pascal Hersen.

Gene expression plays a central role in the orchestration of cellular processes. The use of inducible promoters to change the expression level of a gene from its physiological level has significantly contributed to the understanding of the functioning of regulatory networks. However, from a quantitative point of view, their use is limited to short-term, population-scale studies to average out cell-to-cell variability and gene expression noise and limit the nonpredictable effects of internal feedback loops that may antagonize the inducer action. In this project, in collaboration with the Hersen Lab at MSC (Paris Diderot University), we show that, by implementing an external feedback loop, one can tightly control the expression of a gene over many cell generations with quantitative accuracy. To reach this goal, we developed a platform for real-time, closed-loop control of gene expression in yeast that integrates microscopy for monitoring gene expression at the cell level, microfluidics to manipulate the cells environment, and original software for automated imaging, quantification, and model predictive control. By using an endogenous osmostress responsive promoter and playing with the osmolarity of the cells environment, we show that long-term control can, indeed, be achieved for both time-constant and time-varying target profiles at the population and even the single-cell levels [6]. Importantly, we provide evidence that real-time control can dynamically limit the effects of gene expression stochasticity. We anticipate that our method will be useful to quantitatively probe the dynamic properties of cellular processes and drive complex, synthetically engineered networks.

6.8. Genome Engineering of Mammalian Cells: Targeted and Efficient Integration of Multi-unit Genetic Payloads

Participants: Grégory Batt, Xavier Duportet.

Targeted integration of multi-unit genetic payloads would greatly benefit elucidating complex cellular mechanisms and implementing new functions in mammalian cells. Current technologies are however time-consuming and require tedious post-integration controls. To address this problem, we propose a modular framework to assemble large multi-unit genetic payloads and target their integration into either one or both alleles of a chromosomal locus of choice. To achieve this, we combine in a two-step process the customizable targeting properties of homing endonucleases with the efficiency and specificity of a large serine recombinase. We have demonstrated that an optimized version of BxB1 recombinase allows the targeted integration of large genetic circuits (up to 7 transcription units, 60kb) into a preintegrated landing pad in the AAVS1 locus, with a significant increase in efficiency compared to other site-specific recombination systems (integration in 10% of transfected cells without selection) . By reducing the time and efforts to generate large populations of isogenic stable cell lines adapted to study multi-component genetic systems, our framework is a valuable tool for mammalian synthetic biology and offers great potential for a broad range of biotechnology and therapeutic applications.

6.9. Reifying Global Constraints

Participants: François Fages, Raphaël Martin, Thierry Martinez, Sylvain Soliman.

Global constraints were introduced two decades ago as a means to model some core aspects of combinatorial problems with one single constraint for which an efficient domain filtering algorithm can be provided, possibly using a complete change of representation. However, global constraints are just constraint schemas on which one would like to apply usual constraint operations such as reification, i.e. checking entailment, disentailment and negating the constraint. This is currently not the case in state-of-the-art tools and was not considered in the global constraint catalog until recently. In [20], we propose a general framework for reifying global constraints and apply it to some important constraints of the catalog, such as the cumulative constraint for instance. We show that several global constraints that were believed to be hard to negate can in fact be efficiently negated, and that entailment and disentailment can be efficiently tested. We also point out some new global constraints that are worth studying from this point of view and provide some performance figures obtained with an implementation in Choco.

This scheme is currently used for compiling the Rules2CP constraint modeling language to Choco, and to internalize search in CSPs through constraint reification.

6.10. Railway Time Tabling Optimization

Participants: François Fages, David Fournier, Thierry Martinez, Sylvain Soliman.

Metros are able to generate electricity on a metro line by braking. This energy is immediately available in the third rail and is lost if no metro in the neighbourhood can consume it. It is thus possible to decrease the total energy consumption of a metro line by synchronizing the accelerations and braking of the metros. In [2], [9], we propose a classification of energy optimization timetable problems and we present a model for optimizing energy consumption which does not significantly alter the quality of service, by subtly modifying dwell times. We show however that this optimization problem is NP-hard. We present a hybrid genetic/linear programming algorithm for computing the distribution of braking metros. In this hybridization, the objective function is computed by a linear program and by a heuristic, and the dwell times are modified by a genetic algorithm. On a typical example with real data, the savings exceed 7%. Furthermore, on a benchmark of the literature for a simpler problem, we discuss the results obtained with our genetic algorithm, a tabu search algorithm and the mixed integer linear program used by the authors.

CONVECS Team

6. New Results

6.1. New Formal Languages and their Concurrent Implementations

LNT is a next generation formal description language for asynchronous concurrent systems, which attempts to combine the best features of imperative programming languages and value-passing process algebras. LNT is increasingly used by the CONVECS team for industrial case studies and applications (see § 6.5) and serves also in university courses on concurrency, in particular at ENSIMAG (Grenoble) and at the Saarland University.

6.1.1. Translation from LNT to LOTOS

Participants: Hubert Garavel, Frédéric Lang, Wendelin Serwe.

The LNT2LOTOS, LNT.OPEN, and LPP tools convert LNT code to LOTOS, thus allowing the use of CADP to verify LNT descriptions. These tools have been used successfully for many different systems (see § 6.5 and § 9.1).

In 2012, in addition to 12 bug fixes, the following enhancements have been brought to these tools:

- We improved the ergonomy of the LNT2LOTOS translator by refining certain command-line options and by making some warning messages more user-friendly.
- We optimized the generated LOTOS code of the "disrupt" and "parallel" composition operators, so as to reduce the number of spurious warnings about impossible synchronizations and, more importantly, to meet the subset of LOTOS supported by the CAESAR compiler (static bound on the number of parallel processes).
- We improved the support for LNT programs that contain several modules by allowing the main process to be defined in any module (not only in the main module).
- We added new predefined functions for the generic data types (lists, sorted lists, and sets), and we updated accordingly the reference manual of LNT. The set types are now implemented correctly by avoiding duplicate elements.

6.1.2. Distributed Code Generation for Process Algebras

Participants: Hugues Evrard, Frédéric Lang.

One goal of CONVECS is to build a tool that generates automatically a distributed implementation of a system specified in LNT. This requires a protocol to realize process synchronization. As far as possible, this protocol must itself be distributed, so as to avoid the bottleneck that would inevitably arise if a unique process would have to manage all synchronizations in the system. A particularity of such a protocol is its ability to support *branching synchronizations*, corresponding to situations where a process may offer a choice of synchronizing actions (which themselves may nondeterministically involve several sets of synchronizing processes) instead of a single one. Therefore, a classical barrier protocol is not sufficient and a more elaborate synchronization protocol is needed.

In 2012, we explored the bibliography on synchronization protocols. Among almost twenty references studied, we selected three existing distributed synchronization protocols that seemed appropriate to our problem. In order to validate these protocols, we designed a tool chain that, given a system described as a parallel composition of LNT processes, generates an LNT specification of an implementation of the system (called the *implementation model*), by incorporating the protocol in the specification to realize the synchronizations. We then used CADP to check for livelocks and deadlocks possibly introduced in the implementation model by the protocol (using MCL and EVALUATOR 4.0), and to verify that the implementation model mimicks the behaviour of the system by equivalence checking (using BISIMULATOR).

Among the three protocols considered, we selected the most promising one [57], which is suitable for generalization to implement synchronization vectors (and hence, the generalized parallel composition operator of LNT). Using the methodology mentioned above, we discovered a previously unknown error in this protocol, which leads to deadlocks in certain situations, and we proposed a correction. An article has been submitted to an international conference.

6.1.3. Translation from an Applied Pi-Calculus to LNT

Participants: Radu Mateescu, Gwen Salaün.

The π-calculus is a process algebra defined by Milner, Parrow, and Walker two decades ago for describing concurrent mobile processes. So far, only a few verification tools have been designed for analyzing π-calculus specifications automatically. Our objective is to provide analysis features for π-calculus specifications by reusing the verification technology already available for value-passing process algebras without mobility. Our approach is based on a novel translation from the finite control fragment of π-calculus to LNT. To the best of our knowledge, this is the first π-calculus translation having a standard process algebra as target language.

In this work, we have also extended the original polyadic π-calculus with data-handling features. This results in a general-purpose applied π-calculus, which offers a good level of expressiveness for specifying mobile concurrent systems, and therefore for widening its possible application domains. As language for describing data types and functions, a natural choice was LNT itself: in this way, the data types and functions used in the π-calculus specification can be directly imported into the LNT code produced by translation.

The translation is fully automated by the tool PIC2LNT 2.0. This enables the analysis of applied π-calculus specifications using all verification tools of CADP, in particular the EVALUATOR 4.0 on-the-fly model checker, which evaluates temporal properties involving channel names and data values. PIC2LNT 2.0 was used for teaching mobile concurrency at Saarland University. A paper describing this work was accepted for publication in an international conference [16].

6.1.4. Translation from EB3 to LNT

Participants: Frédéric Lang, Radu Mateescu.

In collaboration with Dimitris Vekris (University Paris-Est Créteil), we have considered a translation from the EB3 language [39] for information systems to LNT. EB3 is inspired from a process algebra, but has the particularity to contain so-called *attribute functions*, whose semantics depend on the history of events. Therefore, the history of events becomes part of the state of an EB3 specification, which is unusual in process algebras.

Since EB3 is not equipped with native verification tools, we have proposed a translation from EB3 to LNT, which would enable EB3 specifications to be formally verified using CADP. Our formal translation scheme ensures the strong equivalence between the LTS corresponding to an EB3 specification and the LTS corresponding to the LNT code generated. The history of events is encoded as a particular LNT process "*memory*" synchronized on all EB3 events with the rest of the system. The memory process thus acts as a monitor that changes its state according to the occurring events and answers requests emitted by the attribute functions when needed. A prototype translator has been developed at University Paris-Est Créteil and a paper describing this work has been submitted to an international conference.

6.1.5. Coverage Analysis for LNT

Participants: Gwen Salaün, Lina Ye.

In the classic verification setting, we have an LNT specification of a system, a set of temporal properties to be verified on the LTS model corresponding to the LNT specification, and a data set of examples (test cases) we use for validation purposes. At this stage, building the set of validation examples and debugging the specification is a complicated task, in particular for non-experts.

Coverage analysis aims at proposing and developing techniques for automatically detecting parts of an LNT specification not (yet) covered during verification. Such LNT coverage analysis techniques would be very helpful for (i) extending the set of test cases with new inputs covering parts of the LNT specification that have not been analyzed yet, (ii) eliminating dead code in the LNT specification, and (iii) extending the set of temporal properties with new ones.

We have already identified four criteria (action, decision, block, property) and developed a prototype tool that automatically returns coverage values for these four criteria. We have applied our tool to LNT specifications of existing protocols, such as a reconfiguration protocol for component-based architectures [34], and found several cases of dead code and missing test cases.

6.1.6. Other Software Developments

Participants: Hubert Garavel, Frédéric Lang, Wendelin Serwe.

In addition of correcting 23 bugs in various CADP tools, we also brought the following enhancements:

- The EUCALYPTUS interface was improved regarding ergonomy and customization.
- The CADP tools for 32-bit and 64-bit Intel/Linux architectures were upgraded to use recent compilers and libraries, and CADP was modified to support Mac OS X 10.8 “Mountain Lion”.
- The usability of the libraries for writing BCG files was improved to detect and signal an improper ordering of the primitives in application programs.
- The SYNTAX parser generator was improved by correcting two subtle errors, one of them causing an infinite looping on certain erroneous input programs. The CADP compilers developed using SYNTAX were enhanced to perform a better diagnosis of the situations when SYNTAX corrected syntactic errors automatically in erroneous programs.
- We improved an optimization of the CAESAR compiler for LOTOS, leading to a significant reduction of the execution time (from one hour and 51 minutes down to 58 seconds) for some examples of LOTOS programs with many variables. We optimized the CAESAR.OPEN script to invoke the CAESAR compiler directly whenever possible (instead of the GENERATOR tool), which improves the performance of graph generation, in particular for LNT.OPEN.
- Four demonstration examples of CADP were extended with LNT descriptions to illustrate the usage of the LNT language and of its compiler. Two examples have been simplified using the latest features of SVL, which can now handle the “n among m” parallel composition operator of LNT. Also, three examples have been reorganized for a better clarity and two couples of examples, which were closely related, have been merged into single examples.

6.2. Parallel and Distributed Verification

Participants: Hubert Garavel, Radu Mateescu, Wendelin Serwe.

For distributed verification, CADP provides the PBG format, which implements the theoretical concept of *Partitioned LTS* introduced in [46] and provides a unified access to an LTS distributed over a set of remote machines. The PBG format is equipped with the DISTRIBUTOR and PBG_MERGE (previously called BCG_MERGE [45]) tools, which perform the distributed generation of a partitioned LTS and the conversion of a partitioned LTS represented in the PBG format into a monolithic LTS stored in a BCG file.

To facilitate the manipulation of partitioned LTSs, CADP provides the PBG_CP, PBG_MV, and PBG_RM tools for copying, moving, and removing PBG files, maintaining consistency during these operations. The PBG_INFO tool provides several functionalities to inspect PBG files, such as checking consistency (i.e., existence and readability of all fragment files), calculating the size (number of states and transitions) of the corresponding LTS, displaying the list of labels, and concatenating remote log files (this is useful, e.g., to understand the reason why a PBG generation fails, and to compute global statistics about CPU and memory usage by the worker processes).

In 2012, in addition to correcting two bugs in DISTRIBUTOR and several bugs in the CAESAR_NETWORK_1 communication library used by the distributed verification tools, we also improved these tools as follows:

- We enhanced DISTRIBUTOR to support more than 256 distributed processes.
- We enhanced CAESAR_NETWORK_1 with a debugging facility, which enables traces of all distributed processes to be generated.
- We enhanced the graphical monitor of DISTRIBUTOR with the option of sorting the labels alphabetically, which facilitates their visual inspection.
- We extended PBG_INFO to enable the display of all labels in a partitioned LTS.

We also developed a prototype tool, named PBG_OPEN, which is an OPEN/CAESAR-compliant compiler for the PBG format, enabling the use of all CADP on-the-fly verification tools on a partitioned LTS. The main advantage of PBG_OPEN is that it can use the memory of several machines to store the transition relation of a partitioned LTS. Therefore, PBG_OPEN can explore on-the-fly large partitioned LTSs that could not be explored using other tool combinations. To reduce the amount of communications, PBG_OPEN can use a cache to store already encountered states, together with their outgoing transitions.

We experimented all these tools on the Grid'5000 computing infrastructure [35] using up to 512 distributed processes. These experiments confirmed the good scalability of our distributed LTS manipulation approach. A paper describing this work has been published in an international conference [12].

6.3. Timed, Probabilistic, and Stochastic Extensions

Participant: Hubert Garavel.

Process calculi provide a suitable formal framework for describing and analyzing concurrent systems, but need to be extended to model refined aspects of these systems. For instance, it may be necessary to represent probabilistic choices (in addition to deterministic and nondeterministic choices) as well as delays and latencies governed by probability laws. Many such extensions have been proposed in the literature, some of which have been implemented in software tools and applied to nontrivial problems. In particular, two of these extensions (namely, *Interactive Markov Chains* and *Interactive Probabilistic Chains*) are implemented in CADP. Despite these achievements, the state of the art is not satisfactory as the extended languages primarily focus on the probabilistic and stochastic aspects, leaving away the expressive and user-friendly features that process calculi provide for describing conventional concurrent systems.

In 2012, we undertook a study to merge probabilistic and stochastic aspects into modern high-level languages such as LNT. This work is done at Saarland University under the aegis of the Alexander von Humboldt foundation, in collaboration also with RWTH Aachen and Oxford University. We investigated the theoretical concepts, as well as their integration into modeling languages, together with the corresponding behavioural equivalences and temporal logics.

We also started experimenting with state-of-the-art software implementations, such as MODEST and PASS (Saarland University), COMPASS and MRMC (RWTH Aachen), and PRISM (Oxford University). Two of these tools (namely, MODEST and PRISM) have been used for lab exercises in the *Applied Concurrency Theory* block course created by H. Garavel at Saarland University. Following these experiments, evaluation reports have been produced, which provide feedback about issues and suggestions for enhancements. These reports have been addressed to the respective authors of each tool and already led to improvements in certain tools.

6.4. Component-Based Architectures for On-the-Fly Verification

6.4.1. Compositional Model Checking

Participants: Frédéric Lang, Radu Mateescu.

We have continued our work on *partial model checking* following the approach proposed in [29]. Given a temporal logic formula φ to be evaluated on a set S of concurrent processes, partial model checking consists in transforming φ into another equivalent formula φ' to be evaluated on a subset of S. Formula φ' is constructed incrementally by choosing one process P in S and incorporating into φ the behavioral information corresponding to P – an operation called *quotienting*. Simplifications must be applied at each step, so as to maintain formulas at a tractable size.

In 2012, we have continued the development of our prototype tools for partial model checking of the regular alternation-free μ-calculus supporting all features of the input language of EXP.OPEN 2.1. We have also extended our work to handle useful fairness operators of alternation depth 2 in linear time, without developing the complex machinery that would be necessary to evaluate general μ-calculus formulas of alternation depth 2. A paper has been published in an international conference [15] and an extended version has been submitted to an international journal.

6.4.2. On-the-Fly Test Generation

Participants: Radu Mateescu, Wendelin Serwe.

In the context of the collaboration with STMicroelectronics (see § 6.5.1 and § 7.1), we studied techniques for testing if a (hardware) implementation is conform to a formal model written in LNT. Our approach is inspired by the theory of conformance testing [59], as implemented for instance in TGV [51] and JTorX [33].

We developed two prototype tools supporting conformance testing. The first tool implements a dedicated OPEN/CAESAR-compliant compiler for the particular asymmetric synchronous product of the model and the test purpose. This tool is a generic component for on-the-fly graph manipulation, taking as input two graphs and producing as output the graph of the asymmetric synchronous product. The second tool generates the complete test graph, which can be used to extract concrete test cases or to drive the test of the implementation. This tool was built from (slightly extended) existing generic components for on-the-fly graph manipulation (τ-compression and τ-confluence reductions, determinization, resolution of Boolean equation systems). The main advantage of our approach compared to existing tools is the use of LNT for test purposes, which facilitates the manipulation of data values.

6.5. Real-Life Applications and Case Studies

6.5.1. ACE Cache Coherency Protocol

Participants: Hubert Garavel, Abderahman Kriouile, Radu Mateescu, Wendelin Serwe.

In the context of a CIFRE convention with STMicroelectronics (see § 7.1), we studied the system-level cache coherency, a major challenge faced in the current system-on-chip architectures. Because of their increasing complexity (mainly due to the significant number of computing units), the validation effort using current simulation-based validation techniques grows exponentially. As an alternative, we study formal verification.

In 2012, we focused on the ACE (*AXI Coherency Extensions*) cache coherency protocol, a system-level coherency protocol proposed by ARM [25]. In a first step, we developed a formal LNT model (about 2600 lines of LNT) of a system consisting of an ACE-compliant cache coherent interconnect, processors, and a main memory. The model is parametric and can be instantiated with different configurations (number of processors, number of cache lines, number of memory lines) and different sets of supported elementary ACE operations, including an abstract operation that represents any other ACE operation. Using the OCIS simulator, we were able to explore the behavior of the system interactively, which has been found helpful by STMicroelectronics engineers.

Currently, our formal model supports a representative subset of five elementary operations of the ACE protocol (MakeUnique, ReadOnce, ReadShared, ReadUnique, and WriteBack). For each of these operations, we have written a liveness property in MCL expressing that the operation is executed until its termination. Using parametric SVL scripts (about 250 lines) and the EVALUATOR 4.0 model checker, we verified these properties on the fly for up to three memory lines and two processors with two cache lines each. We also generated the corresponding LTS (up to 250 million states and one billion transitions).

We also started considering data integrity properties. This required to translate a state-based property (namely, the consistency between the values stored in memory and in the local caches of the processors) into our action-based setting. This enabled us to automatically exhibit a known error present in a previous version of the ARM specification of the ACE protocol (which was corrected in a subsequent version of the specification). Using the LNT model corresponding to the latest version of the ACE specification, we spotted several potential data integrity issues that we reported to STMicroelectronics, where they are currently under investigation.

6.5.2. Realizability of Choreographies

Participants: Alexandre Dumont, Matthias Güdemann, Gwen Salaün.

Choreographies allow business and service architects to specify, with a global perspective, the requirements of applications built over distributed and interacting software entities. In collaboration with Pascal Poizat (University Paris-Sud), we proposed new techniques for verifying BPMN 2.0 choreographies, and particularly the *realizability* property. Realizability ensures that peers obtained via projection from a choreography interact as prescribed in the choreography requirements. Our approach is formally grounded on a model transformation into the LNT process algebra and the use of equivalence checking. It is also completely tool-supported through interaction with the Eclipse BPMN2 modeler and CADP. These results have been published in an international conference [17].

In collaboration with Meriem Ouederni (University of Nantes), we extended our techniques for analyzing choreographies to restore realizability for non-realizable, but *repairable* choreographies. For this we exploit the counterexamples generated by the equivalence checker BISIMULATOR to identify problematic messages in the choreography. For those messages we add distributed local monitors to the system which delay message sending if necessary, to restore correct message sequences. This iterative approach introduces the minimal number of necessary additional messages to restore realizability, and the monitors are generated in the most permissive way, i.e., by considering all possible interleavings given the behaviour of the peers participating to the choreography. It is fully automated by a prototype tool we implemented. These results have been published in an international conference [14].

We developed a common formal description language, named CIF (*Choreography Intermediate Format*), for the verification of choreographies. CIF is based on an XML representation for easy exchange between programs, an XSD schema for validation, and a translation to LNT for verification. CIF is used as an intermediate language to specify choreographies, but can also serve as target language for translating various choreography specification languages, such as BPMN 2.0. The back-end connection to CADP via LNT enables the automation of some key choreography analysis tasks (repairability, realizability, conformance, etc.). Our framework is extensible with other front-end and back-end connections to, respectively, other choreography languages and formal verification tools.

6.5.3. Self-Configuration Protocol for Distributed Cloud Applications

Participants: Rim Abid, Gwen Salaün.

We collaborate with Noël de Palma and Fabienne Boyer (University Joseph Fourier), Xavier Etchevers and Thierry Coupaye (Orange Labs) in the field of cloud computing applications, which are complex, distributed artifacts involving multiple software components running on separate virtual machines. Setting up, (re)configuring, and monitoring these applications are complicated tasks because a software application may depend on several remote software and virtual machine configurations. These management tasks involve many complex protocols, which fully automate these tasks while preserving application consistency as well as some key properties.

In this work, we focus on a self-configuration protocol, which is able to configure a whole distributed application without requiring any centralized server. The high degree of parallelism involved in this protocol makes its design complicated and error-prone. In order to check that this protocol works as expected, we specify it in LNT and verify it using the CADP toolbox. The use of these formal techniques and tools helped to detect a bug in the protocol, and served as a workbench to experiment with several possible communication models. These results led to a publication in an international conference [18].

We are currently studying two variants of the self-configuration protocol, one handling virtual machine failures, and one allowing dynamicity in the system (addition and removal of virtual machines) using a publish-subscribe communication framework.

6.5.4. Networks of Programmable Logic Controllers

Participants: Hubert Garavel, Fatma Jebali, Jingyan Jourdan-Lu, Frédéric Lang, Eric Léo, Radu Mateescu.

In the context of the Bluesky project (see § 8.1.2.1), we study the software applications embedded on the PLCs (*Programmable Logic Controllers*) manufactured by Crouzet Automatismes. One of the objectives of Bluesky is to enable the rigorous design of complex control applications running on several PLCs connected by a network. Such applications are instances of GALS (*Globally Asynchronous, Locally Synchronous*) systems composed of several synchronous automata embedded on individual PLCs, which interact asynchronously by exchanging messages. A formal analysis of these systems can be naturally achieved by using the formal languages and verification techniques developed in the field of asynchronous concurrency.

For describing the applications embedded on individual PLCs, Crouzet provides a dataflow language with graphical syntax and synchronous semantics, equipped with an ergonomic user interface that facilitates the learning and use of the language by non-experts. To equip the PLC language of Crouzet with functionalities for automated verification, the solution adopted in Bluesky was to translate it into a pivot language (to be defined within the project) that will enable the connection to testing and verification tools covering the synchronous and asynchronous aspects. Our work focuses on the translation from the pivot language to LNT, which will provide a direct connection to all verification functionalities of CADP, namely model checking and equivalence checking.

In 2012, in interaction with Crouzet engineers, we studied the PLC language of Crouzet to understand precisely its static and dynamic semantics. We specified manually in LNT several examples of control applications provided by Crouzet, with the goal of identifying the principles of translating the PLC language of Crouzet to LNT. We formulated in MCL several safety and liveness properties concerning the temporal ordering of input and output events by the control applications, and we successfully verified them on the LNT specifications. We also started to study the network communication mechanisms between PLCs to identify a suitable LNT abstraction of the communication layer.

6.5.5. Other Case Studies

Participants: Frédéric Lang, Radu Mateescu, Wendelin Serwe.

Continuing a study [53] started in the context of the Multival project (see http://vasy.inria.fr/multival), we considered the Platform 2012 architecture proposed by STMicroelectronics, focusing on the Dynamic Task Dispatcher (DTD), a hardware block that assigns a set of application tasks to a set of processors. In 2012, we extended our LNT model and the corresponding MCL properties in order to handle heterogeneous processors equipped with different kinds of processor extensions. We also used constraints on the initialization phase, which reduced the size of the LTS by a factor of up to ten and hence enabled the generation of the LTS for up to eight processors (instead of only six). Both extensions together enabled to discover the possibility of a livelock.

We attempted to investigate this issue further by cosimulation (using the EXEC/CAESAR framework) with the original C++ model of the architect. Unfortunately, the C++ model did not behave correctly for the particular aforementioned application scenario. It was not possible to change this model because the recent evolutions of Platform 2012 excluded the DTD, as its requirements in terms of silicon surface were considered too large. This work, including the LNT model as appendix, has been accepted for publication in an international journal [5].

In collaboration with Nuno Mendes and Claudine Chaouiya (Gulbenkian Institute, Portugal), Yves-Stan Le Cornec (IBISC, University Evry Val d'Essonne) and Grégory Batt (CONTRAINTES project-team, Inria Paris-Rocquencourt), we have studied the use of CADP for checking the reachability of stable states in genetic regulatory networks. A compositional and logical model of genetic regulatory networks called *logical regulatory modules* was defined and translated to LNT processes and EXP.OPEN 2.1 networks of LTSs.

Compositional minimization modulo safety equivalence was applied to the generated network, so as to palliate state explosion while preserving the reachability property. The approach has been illustrated on the segment polarity module involved in the segmentation of the fruit fly embryo and on the delta-notch module involved in cell differentiation in crucial steps of embryonic development of several species. A paper has been submitted to an international journal.

DART Project-Team

6. New Results

6.1. Hardware Distributed Control for Dynamic Reconfigurable Systems

The progress in FPGA technology has allowed FPGA-based reconfigurable embedded systems to target increasingly sophisticated applications, which leads to a high design complexity of such systems especially at the adaptation control level. This complexity results into long design phases and delayed time-to-market. In this context, a centralized control model might be not adapted to the growing size and complexity of embedded systems. The use of a single controller for the whole system might result into a high complexity due to the number of parameters to take into account for runtime adaptation, which makes difficult its modification and test. Besides, the design of such a controller is system-dependent since it treats the system as a whole, which represents an obstacle for design reuse. In order to solve these problems, we propose a control design approach aiming to decrease design complexity and enhance design flexibility, reuse and productivity. This approach is based on a semi-distributed control model [34]. In order to achieve the objectives mentioned above, the proposed approach combines autonomy, modularity, formalism and high-level design. The semi-distributed control model divides the control problem between autonomous controllers handling each the self-adaptation of a reconfigurable component of the system, which allows to decrease their design complexity. Each controller handles three main tasks allocated to three different modules: i)monitoring of events that might trigger the adaptation of the controlled component, ii)decision-making about the required adaptations, and iii)adaptation (reconfiguration) realization. To ensure that reconfiguration decisions made by the controllers respect global system constraints such as security and quality of service constraints, these decisions are coordinated before launching the corresponding partial reconfigurations. The allocation of these tasks to separate modules facilitates their modification and reuse and thus the scalability of the control design. For the decision-making modeling, we use the mode-automata formalism. This formalism is suitable to model the control of the different modes of a reconfigurable system such as energy modes or image display modes. Thanks to its clear semantics, the use of such a formalism facilitates the high-level modeling of the controllers and their automatic generation. In order to facilitate code generation and enhance thus design productivity, our control approach makes use of Model-Driven-Engineering (MDE) [33]. Control systems composed of controllers and coordinators are modeled using the UML (Unified Modeling Language) profile MARTE (Modeling and Analysis of Real-Time and Embedded systems). The automation of MDE, allowed to generate the code of these systems. The generated code was then used to validate the semi-distributed control and to determine its resource overhead compared to centralized control systems.

6.2. Regular interconnection network for HP-SoC architecture

Our Synchronous Communication Asynchronous Computation (SCAC) model is a data-parallel execution model dedicated to the High Performance System-on-Chip. The architecture of this model is composed of huge number of complex routers, called node elements (the NEs), communicating and working in perfect synchronizations. Each NE is potentially connected to its neighbors via a regular connection. Furthermore, each NE is connected to a heterogeneous set of computing groups (clusters) allow asynchronous processing. Each group includes a combination of processors programmable, the PEs (software processing units) and specialized hardware accelerators (hardware processing units) to perform critical tasks demanding the more performance. All the system is controlled by a Network Controller Unit, the NCU. The NCU and The PEs are implemented with the Forth processor.

The synchronous communication in SCAC model is presented by two kinds of communications:

- The NCU/NEs communication. In fact, we defined a hNoC model integrated in the SCAC architecture [31]. This model is based on sub-netting the network of processing nodes which separate the control of communication and processing. From this model, our communication system allows a better management of data congestion in the NEs grid through the broadcast with mask of parallel instructions to activated processing nodes.
- The NE/NE communication which is our last contribution. In fact, we defined the X-net interconnection network which is a regular network dedicated to the massively parallel SCAC architecture. This network interconnects directly each PE with its 8 nearest neighbors in a two-dimensional mesh through a specific router in the NE module.

The aim of these last works is to design a regular NoC for SCAC architecture to allow global synchronization of the system communications and increase high performance in terms of area cost and bandwidth. This network based on IP blocks which offer well flexibility and scalability, was implemented in synthesizable VHDL code that was simulated and targeted Xilinx Virtex6 (XC6VLX240T) board. The difficulty of designing X-net is a compromise between an optimal quality of broadcasting, high bandwidth and important flexibility of use, while reducing power consumption and silicon area.

6.3. ReCoMARTE: A Marte Based Profile for Dynamic Reconfigurable Systems Modeling

During the last decade, DPR has been widely studied as a research topic. Despite its intuitive appeal, the technique had eluded widespread adoption, particularly in industrial applications. This is due to the complexities of the provided design flow and the in-depth knowledge of many low level aspects of FPGA technologies used to implement DPR systems. The aim of our current work is to propose a methodology in order to allow us to introduce PDR in MARTE for modeling all types of FPGAs supporting our chosen PDR flow. Afterwards, using the MDE model transformations, the design flow can be used to bridge the gap between high level specifications and low implementation details to finally generate files used by the Xilinx EDK design flow for implementing the top-level SoC description of the system. Indeed, in its current version, UML MARTE profile lacks dynamic reconfiguration concepts and requirements for the reconfiguration at different abstraction levels. We have concentrated our efforts in the creation of the structural description of the system that is used as an input to the DPR design flow to facilitate the design entry phase of the DPR design flow. Therefore, we defined an extended version of MARTE called RecoMARTE (Reconfigurable MARTE) [16] model these concepts mainly at:

- Application level: For reconfigurable applications combining control and data processing, it is very difficult, even impossible to use the MARTE profile for their specification. Non-functional properties such as control concepts are induced by different configurations or running modes of the system and allow the description of more complex behaviours. We recommend a set of extensions to a MARTE profile. We also focus on modelling heterogeneous reconfigurable components, and address the problem of constraints specification for verification issue.
- Control mechanism: We define necessary requirements for the reconfiguration control mechanism in order to manage reconfiguration at every design level. In addition, our solution allows to describe global contracts and constraints for combining automata. Our modeled reconfiguration controller will be then synthesized using Discrete Controller Synthesis formal technique (collaboration work)I n order to always provide a correct configuration to the system, with respect to constraints specified by the designer
- Deployment level: Our design methodology using RecoMARTE enables the deployment, parameterization and integration of hardware IPs into SoC platform at multiple levels of abstraction. We have introduced IP deployment capabilities in MARTE, which aim at facilitating the import of selected low-level features into the high-level models, their modification, and the creation of an IP-XACT design description that is used to parameterize and integrate the underlying IP descriptions.

- Physical level: introduced extensions in MARTE provide some facilities to allow modeling physical architecture of a chosen FPGA. Our solution allows to carry out the physical placement of static and reconfigurable areas on the platform. This task is done through ranges in terms of physical resources, with respect to placement constraints such as consumed resources.

6.4. Using Marte Profile for NoCs modeling

The modeling of repetitive structures such as network on chip topologies in graphics forms poses a particular challenge. This aspect may be encountered in intensive data/control oriented applications such as H.264 video coder. In this work we have described an adequate methodology for modeling NoCs by using the MARTE standard profile. The proposed study has shown that the Repetitive Structure Modeling (RSM) package of MARTE profile is powerful enough for modeling different topologies. By using this methodology, several aspects such as routing algorithm are modeled based finite state machines. This allows to the MARTE profile to be complete enough for modeling a large number of NoCs architectures. Some work is on-going to synthesize such networks in VHDL from such models [55]. While validating the proposed methodology, a co-design approach has been studied by mapping a H264 video coding system onto a Diagonal Mesh NoC by using the Y Chart of Gaspard2 tool. Before allowing the association of the application/architecture, an architectural optimization targeting power minimization of the most critical module of the application and the router of the architecture has been performed. For instance, a flexible VLSI architecture for full-search VBSME (FSVBSME) has been proposed.

6.5. A Hardware Membranes Based Reconfiguration Services Implementation

Partial and dynamic reconfiguration provides a relevant new dimension to design efficient parallel embedded systems. However, due to the encasing complexity of such systems, ensuring the consistency and parallelism management at runtime is still a key challenge. So architecture models and design methodology are required to allow for efficient component reuse and hardware reconfiguration management. We proposed a novel approach inspired from the well-known component based models used in software applications development. Our model is based on membranes wrapping the systems components. The objective is to improve design productivity and ensure consistency by managing context switching and storage using modular distributed hardware controllers. These membranes are distributed and optimized with the aim to design self-adaptive systems by allowing dynamic changes in parallelism degree and contexts migration [26]. These results are obtained in the Famous project by a collaboration with LABSticc Lorient.

6.6. Formal Techniques for General and Domain-Specific Languages

In 2012 we have finished the previous year's activities on domain-specific languages based on formal model-driven engineering with two papers [18], [24]. Our conclusion is that formal MDE-based language definition is interesting because of its generality but adds extra layers of complexity due to the fact that language concepts and semantics are only formalised indirectly, through the formalisation of MDE concepts used in language definition. We have decided thus to move on towards more direct ways of defining and reasoning about languages. We have been experimenting with the K framework [4] for formally defining both the assembly language and a higher-level language for programming on the upcoming dynamically reconfigurable hardware architecture that our team is developing.

We have also worked on proving the correctness of a compiler between high-level and assembly language, based on new symbolic program-equivalence proof techniques we are developping in collaboration with the K team [29]. We have also been working on generic symbolic execution techniques for programming languages having term-rewriting based semantics [28] (PhD of Andrei Arusoaie, supervised in collaboration with Prof. Dorel Lucanu from the K team of Univ. Iasi (Romania).

[4] http://www.k-framework.org

DEDUCTEAM Team

5. New Results

5.1. Dedukti

Together with Mathieu Boespflug (McGill University), Quentin Carbonneaux and Ronan Saillard have developed a new version of the front-end of Dedukti, written in OCaml, replacing an inefficient previous version, as well as a new version of the back-end using the Lua Just-In-Time compiler.

Ronan Saillard has internalized the Lua back-end of Dedukti, so that it is no longer necessary to explicitly call it when using Dedukti.

Ronan Saillard has extended the input language of Dedukti to allow the user to declare dependencies between modules, to write definition or to explicitly require to type-check a term.

Ronan Saillard has added a new feature to Dedukti to make opaque definitions. As with usual definitions, the proof term of an opaque definition is type-checked, but it is then immediately forgotten in order to decrease memory consumption.

5.2. Embeddings in the $\lambda\Pi$-calculus modulo

Ali Assaf has designed an embedding of the HOL logic in the $\lambda\Pi$-calculus modulo and implemented it in the HOLiDe system [40].

Together with Mathieu Boespflug, Ali Assaf and Guillaume Burel have developed an embedding of the Calculus of Inductive Constructions with universes in the $\lambda\Pi$-calculus modulo and Ali Assaf is currently implementing it in a new version of the CoqInE system.

Catherine Dubois and Raphaël Cauderlier have studied a translation in the $\lambda\Pi$-calculus modulo of features coming from object oriented programming languages, such as inheritance and late binding. This compilation scheme has been applied to produce a new back-end for FoCaLize [8], through a compilation to Dedukti. This new back-end is expected to be lighter than the present one producing Coq code and also to be able to combine local and external proofs coming from different proof environments [44]. They are currently working on a translation of the full FoCaLize language—not restricted to its object oriented features—and on a proof of its correctness with respect to the existing FoCaLize semantics.

5.3. Automated Theorem Proving

Guillaume Burel has shown that presenting theories by means of rewriting rules in Deduction modulo leads to more efficient proof search methods than using axioms, provided the rewriting system enjoys a proof theoretical property, namely cut admissibility.

He has been investigating which theories can be encoded as rewriting systems admitting cuts. Surprisingly, it turned out that any consistent theory in predicate logic can. This has been shown by studying the links between the set-of-support strategy of the Resolution method and the extension of the method based on Deduction modulo. He has also shown how to reduce the size of the corresponding rewriting systems [42].

Guillaume Burel has also studied how to improve the confidence in iProver Modulo. When it finds a resolution proof, it is now able to produce a proof that can be checked by Dedukti. The encoding of Resolution proofs in the $\lambda\Pi$-calculus modulo that is used is shallow, making more plausible the long-term goal of interoperability of provers, both interactive and automated, through Dedukti.

Simon Cruanes has explored several ideas for combining the Superposition calculus—one of the most powerful calculi for automated reasoning within first-order logic with equality—with Deduction modulo. Combining the term rewriting system for a theory in Deduction modulo with the ordered rewriting on which Superposition is based on proved to be difficult, yielding incomplete calculi; in most cases it boils down to the fact that the combination of confluent terminating term rewriting systems is in general neither terminating nor confluent. In order to experiment quickly ideas by implementing them, he has written a Superposition-based prover in OCaml, with some special features—automatic ordering of rewrite rules in the input, non-clausal calculus to be able to use equivalence relations as rewrite rules. The prover is 8,000 lines of code and is designed to be flexible and modular, but still has decent performance and can prove some non-trivial theorems.

Together with Mélanie Jacquel (Cedric), David Delahaye and Catherine Dubois have investigated Zenon for verifying proof rules added to help the automation in the provers of Atelier B. They have augmented Zenon with specific rules for dealing with set operations and predicates, obtained by applying super deduction—a variant of Deduction modulo [33].

5.4. Proof theory

We believe that our work on proof-checking and automated theorem proving cannot be separated from a more theoretical research on proof theory.

Together with Denis Cousineau, Gilles Dowek and Olivier Hermant have related semantic criteria for proof normalization and admissibility of the cut rule in Deduction modulo [17], [26].

Gilles Dowek has proposed a new way to define classical connectives in a constructive framework [46].

Together with Murdoch J. Gabbay (Heriot Watt), Gilles Dowek has proposed a new nominal logic that handles binders in terms [16] and a new semantics for predicate logic [29].

During her visit in the team, Cecilia Englander has studied the correspondence between natural deduction and sequent calculus.

Together with Ying Jiang (Beijing), Gilles Dowek has defined a logic for finite structures. Kailiang Ji is currently investigating the use of proof search algorithms in Deduction modulo to automatically prove theorems in this theory.

5.5. Safety of aerospace systems

Together with Anthony Narkawicz (Nasa-Langley) and César Muñoz (Nasa-Langley), Gilles Dowek has designed a prevention bands algorithm, that is an algorithm that computes and displays to the pilot of an aircraft, a sequence of safe and unsafe intervals on ground speed, heading or vertical speed and they have proved this algorithm correct in the PVS system [18].

This algorithm computes with real numbers, but its implementation computes with floating point numbers. Moreover this algorithm is numerically unstable as it uses comparisons of numbers, computed with square root and division operations. This has led Pierre Néron to design a program transformation algorithm to eliminate square roots and divisions in straight-line programs. This way computation can be made exact.

Together with César Muñoz, Pierre Néron has completed this year the design of this program transformation algorithm and he has proved, in the PVS system, its termination and correctness: preservation of semantics and absence of square roots and divisions in the produced program [35].

Together with César Muñoz, Pierre Néron has also implemented this transformation algorithm as a PVS automatic proof strategy, that allows a wider range of expressions, using a deep embedding of PVS in PVS itself.

Pierre Néron and Raphaël Bost have proposed an optimization of one aspect of that algorithm: the definition of a common template for arithmetic expression.

5.6. Constraint Solving

Catherine Dubois has developed in collaboration with Matthieu Carlier and Arnaud Gotlieb (Oslo) a formally verified constraint finite domain solver. It focuses on arc-consistency and has been developed with Coq [24].

5.7. Models of Computation

Together with Pablo Arrighi (Grenoble), Gilles Dowek has reformulated Gandy's proof of the physical Church-Thesis in the quantum case [11]. Gilles Dowek has proposed the idea that the Galileo thesis could be seen as a consequence of the physical Church-Turing thesis and therefore as a consequence of Gandy's principles [15]. Gilles Dowek has proposed a definition of a notion of non deterministic computation over the real numbers [14] that could be used as a language to describe continuous non deterministic physical phenomena. All this work has then been presented in a tutorial at the conference *Language and Automata Theory and Applications* [28].

Together with Pablo Arrighi, Gilles Dowek has investigated further the principle of a finite density of information [38] and in particular the impact of this definition on the notion of a chaotic dynamical system [37].

Together with Pablo Arrighi, Gilles Dowek has investigated a generalization of the notion of cellular automaton where the principle of a bounded density of information is formulated independently of the geometry of space. This led to the notion of a Causal graph dynamic [12].

Nachum Dershowitz and Gilles Dowek have shown that extending Turing machines with a two-dimensional tape, made this formalism usable in practice to implement classical algorithms [45].

Alejandro Díaz-Caro and Gilles Dowek have proposed to take a fresh look at non deterministic λ-calculi—such as quantum λ-calculi—and derive non determinism from type isomorphism [30].

Together with Giulio Manzonetto (Paris 13) and Michele Pagani (Paris 13), Alejandro Díaz-Caro has considered an extension of the call-by-value λ-calculus with a may-convergent non-deterministic choice and a must-convergent parallel composition, endowed with a type system. They have proved that a term is typable if and only if it is converging, and that its typing tree carries enough information to give a bound on the length of its lazy call-by-value reduction. Moreover, when the typing tree is minimal, such a bound becomes the exact length of the reduction [31].

Together with Barbara Petit (Sardes), Alejandro Díaz-Caro has considered the non-deterministic extension of the call-by-value lambda calculus, which corresponds to the additive fragment of the linear-algebraic lambda-calculus. They have defined a fine-grained type system, capturing the right linearity present in such formalisms. After proving the subject reduction and the strong normalisation properties, they have proposed a translation of this calculus into the System F with pairs, which corresponds to a non linear fragment of linear logic. The translation provides a deeper understanding of the linearity in this setting [32].

Together with Pablo Arrighi, Barbara Petit, Pablo Burias (Rosario), Mauro Jaskelioff (Rosario), and Benoît Valiron (Penn), Alejandro Díaz-Caro has studied possible typing systems for the full linear-algebraic λ-calculus in which the non-deterministic calculus can be seen as a particular case. They have proposed a type system that keeps track of "the amount of a type" that is present in each term [13]. As an example of its use, they have shown that it can serve as a guarantee that the normal form of a term is barycentric, that is that its scalars are summing to one. They also proposed a type system similar to the one presented in [32], but for the full calculus, ensuring confluence and convergence [23]. Finally, they provided a full type system that is able to statically describe the linear combinations of terms resulting from the reduction of programs, also ensuring convergence [19].

ESPRESSO Project-Team

6. New Results

6.1. Extensions of the Signal language and the Polychrony formal model

Participants: Thierry Gautier, Paul Le Guernic.

The different works related to the use of the polychronous model as semantic median model (which has also a syntactic instance) for different effective models (AADL [15], Simulink via GeneAuto, UML via CCSL...) lead us to study various possible extensions of the semantic model as well as the syntactic one.

Thus, we are defining a new version, V5, of Signal, that will be a deep evolution from the current V4 version.

In particular, we are investigating the way state diagrams are best represented in the polychronous model of computation, maintaining the multi-clock characteristic property of the representation. We propose a semantic model for these automata, that relies on the Boolean algebra of clocks. A special case of automata is those that may be represented as regular clock expressions, for which we develop a specific formal calculus. These regular expressions may be used as a powerful manner to express regular dynamic properties of polychronous processes. In correspondence with these models, we are defining syntactic structures to represent these Signal State Diagrams.

Moreover, an important challenge we want to address in the next few years is that of providing design automation techniques and tools for engineering heterogeneous cyber-physical systems (CPS). This leads in particular to new requirements related to the language itself in which we want to describe such software architectures. With respect to the current V4 version of Signal, the basic idea is to extend Signal with a syntactic structure that encapsulates a polychronous process P in a *system*, S, that could have a continuous temporal domain providing a real-time clock presented in some time unit (*fs*, ..., *ms*, ..., *sec*, *mn*, ...). Such a real-time clock can be used as a usual signal in the process P encapsulated in S. Systems S_1, ..., S_n may be composed (with the standard composition of Signal) in a same system S, but the *ms* of a given system S_i is a priori not synchronous with the *ms* of another system S_j. Then it is possible to specify constraints in the system S on these different signals, to express for instance some variation limits of different clocks.

For that purpose, we have defined a new taxonomy of polychronous processes to characterize precisely the following classes: system, task, (logical) process, function, reaction, diagram, observer, controller... This characterization is based on properties such as time reference, input-output clock relations, input-output dependences, determinism, exo/endochrony. For example, a system is either a physical system abstraction, or a basic system, or a system of systems. A basic system has a unique continuous time reference; it provides an internal actual discrete time unit subset of its external continuous time, shared by all its components. As another example, for a subclass of logical processes: a function is a deterministic, inout clocked, endochronous and atomic process that denotes a mathematical flow function. All these different semantic classes are provided syntactic counterparts in the new Signal V5.

6.2. Experimental Polarsys platform

Participants: Loïc Besnard, Thierry Gautier, Jean-Pierre Talpin.

In the context of the OPEES project (http://www.opees.org/), we have experimented the IWG Eclipse platform Polarsys (http://www.eclipse.org/org/press-release/20111102_polarsys.php). Polarsys is a new industry collaboration to build open source tools for safety-critical software development. The integration of Polychrony into this platform has been realized in collaboration with the CS company. CS and Inria have produced the Polychrony experimentation report which is included in the global experimentation report. This document gathers the experiments performed by the several partners involved in the OPEES project on the Polarsys platform. An experiment is defined as the way one partner takes his component and uses it to check any of the services within the Polarsys environment. The services are functions the partners want the Polarsys environment to offer.

For the qualification of the Polychrony component on the Polarsys platform, CS and Inria provide the following documents:

- The Tool Quality Assurance Plan Template (TQAP). This document defines the OPEES quality assurance arrangements and gives some guidance to satisfy them. It focuses on qualification aspects and gives in appendices guidance for some criteria tool qualification with an example for Polychrony Tool.
- The Tool Verification Cases and Procedures (TVCP) document. It presents the test cases to be performed for the qualification of Polarsys Polychrony Verifier component as described in the TQAP.
- The Tool Verification Results (TVR). It presents the results of tests performed for the qualification of Polychrony Verifier on several operating systems, as described in the TQAP.

6.3. Translation validation of Polychronous Equations with an iLTS Model-checker

Participants: Van-Chan Ngo, Jean-Pierre Talpin, Thierry Gautier, Paul Le Guernic, Loïc Besnard.

This work [16], [18], which is part of the VeriSync project, focuses on verifying the correctness of transformations on abstract clocks in the Signal compiler [8]. We propose to use model checking technique over Polynomial Dynamical Systems (PDS) with the Sigali model checker [39].

Adopting the *translation validation* approach of [55], [54], we present an automated verification process to prove the correctness of a multi-clocked synchronous language compiler. Due to the very important role of abstract clocks and clock relations, we are interested in proving that abstract clocks and clock relations semantics of source programs are preserved during the compilation phases. Each individual transformation or optimization step of the compiler is followed by our verification process which proves the correctness of this step. The compiler will continue its work if and only if the correctness is proved positively. This approach avoids the disadvantage of proving in advance that the compiler always behaves correctly since every small change to the compiler requires reproving it.

Our verification framework uses polynomial dynamical systems (PDS) over a finite field, as common semantics for both source and transformed programs. We formalize the abstract clocks semantics of *polychronous equations* with the finite field modulo $p = 3$ as a PDS [16]. For a signal x, if x is boolean, we use the values $-1, 0, +1$ to encode (respectively) the fact that it is present and false, absent, or present and true. Then, the abstract clock can be represented by x^2. If x is non-boolean, we only encode the abstract clock by x^2, meaning that $x^2 = 0$ encodes x is absent, $x^2 = 1$ encodes x is present. An appropriate relation called *refinement* for PDSs is proposed to represent the correct transformation relation between the source and transformed programs. Then a dedicated checking procedure is proposed within Sigali to check the correct transformation relation. The checking procedure is based on the simulation techniques [30]. It is implemented as extension function of the Sigali model checker within the Polychrony toolset.

We have proposed an approach to prove the clock semantic preservation of the Signal compiler transformations until the generation code phase as well. The verification method applied to code generation phase addresses the formal verification of the generated C-code from a refined and optimized intermediate specification in which the compiler enforces logical timing constraints and in which the execution order of data-flow equations is completely scheduled. As a result, all individual transformations, optimizations, and code generation phases of the compiler are followed by a verification step which proves the correctness of transformations. The compiler continues if and only if correctness is proved and returns an error and a trace otherwise. The main idea is that the sequential C code is translated into the target synchronous program thanks to the intermediate SSA form, which is based on the work in [3]. In addition, if a refinement relation between two PDSs does not exist, our validator will find the set of states along with their associated events, which can be used to construct counterexamples in the transformed program [18].

6.4. Formal Verification of Transformations on Abstract Clock in Synchronous Compilers

Participants: Van-Chan Ngo, Jean-Pierre Talpin, Paul Le Guernic.

Translation validation was introduced in the 90's by Pnueli et al. as a technique to formally verify the correctness of code generated from the data-flow synchronous language Signal using model checking. Rather than certifying the code generator (by writing it entirely using a theorem prover) or qualifying it (by obeying to the 27 documentations required as per the DO-178C), translation validation provides a scalable approach to assessing the functional correctness of automatically generated code. By revisiting translation validation, which in the 90's suffered from the limitations of theorem proving and model checking techniques available then, we aim at developing a scalable and flexible approach that applies to the existing 500k-lines implementation of Signal, Polychrony, and is capable of handling large-scale, possibly automatically generated, Signal programs, while using of-the-shelf, efficient, model-checkers and SAT/SMT-solving libraries [36], [63].

The abstract clock semantics of polychronous equations is formalized as a first-order logic formula over boolean variables. For a signal x, if x is boolean then we use two boolean variables x, and $\widehat{x}$ to represent the value of signal x and it abstract clock, respectively. If x is non-boolean signal, we only need to capture its abstract clock by a boolean variable $\widehat{x}$. The boolean variable $\widehat{x}$ is true when the signal x is present and otherwise it is absent. The equational structure of a synchronous data-flow language makes that it is natural to represent the relations between abstract clocks described implicitly or explicitly by equations in terms of first-order logic formulas. And then the combination of equations can be represented by the conjunction of the corresponding first-order logic formulas. We assume that all considered programs are supposed to be written with the primitive operators, meaning that derived operators are replaced by their corresponding primitive ones, and there is no nested operators such as $z := x$ default $(y$ when $b)$. The nested operators are broken by using fresh variables. These formulas use the usual logic operators and numerical comparison functions. Consider a general equation $y := R(x_1, x_2, ..., x_n)$, where R is an operator defined in a synchronous language (e.g. `suspend` in Esterel, `when` in Signal...), or it can be a usual boolean or numerical operator, then the abstract clock semantics of this equation can be represented as a first-order logic formula over the clocks and/or the boolean value of the involved signals $\Phi(\widehat{y}, \widehat{x_1}, \widehat{x_2}, ..., \widehat{x_n}, x_1, ...)$. For a boolean expression defined by numerical comparison functions and numerical expressions, to ensure the result formulas are boolean, we only encode the fact that the clocks of boolean and numerical expressions are synchronized, and we avoid encoding the numerical comparison function which defines the value of the boolean expression and the numerical expressions. For each i^{th} equation in program P, we denote by Φ_{eq_i} its abstract clock semantics, then the abstract clock semantics of P can be represented by a first-order logic formula, called its *clock model*, denoted as:

$$\Phi_P = \bigwedge_{i}^{n} \Phi_{eq_i} \tag{1}$$

where n denotes the number of equations composed in P. The detailed encoding scheme of the Signal language can be found in [19].

Given two clock models P_1 and P_2, we say that P_2 is a *correct transformation* on abstract clocks of P_1, or P_2*refines*P_1 w.r.t the clock semantics, if it satisfies:

$$\forall I.(I \models \Phi_{P_2} \rightarrow I \models \Phi_{P_1}) \tag{2}$$

We write $P_2 \sqsubseteq_{clock} P_1$ to denote the fact that P_2 refines P_1. We also provide an approach to check the existence of refinement by using a SMT-solver that is based on the following theorem:

Theorem. Given a source program P_1 and its transformed program P_2, P_2 is a correct transformation of P_1 on abstract clocks if it satisfies that the formula Φ_{P_1} is a logical consequence of the formula Φ_{P_2}, or

$$\models(\Phi_{P_2} \rightarrow \Phi_{P_1}) \text{ if and only if } P_2 \sqsubseteq_{clock} P_1 \tag{3}$$

Here, we delegate the checking of the above formula against the clock models to a SMT-solver that efficiently deals with first-order logic formulas over boolean and numeric expressions. The checking formulas belong to decidable theories, this solver gives two types of answers: *sat* when the formula has a model (there exists an interpretation that satisfies it); or *unsat* otherwise. Our implementation uses the SMTLIB common format [31] to encode the formulas obtained from the previous step as input of SMT solvers. For our implementation, we consider the Yices solver [38], which is one of the best two solvers at the last SMTCOMP competition [59].

6.5. Formal Verification of Transformations on Data Dependency in Synchronous Compilers

Participants: Van-Chan Ngo, Jean-Pierre Talpin, Paul Le Guernic.

We propose an approach to prove the data dependency semantic preservation of transformations in a synchronous compiler (such as that of Signal). In the Signal language, the scheduling or data dependency is expressed implicitly through polychronous equations. We use *Synchronous Data-flow Dependence Graphs* (SDD Graphs) [46], [50] to formalize the data dependency semantics of polychronous equations. A tuple $< G, C, fE >$ is a SDD graph if and only if:

- $G =< N, E, I, O >$ is a dependence graph $< N, E >$ with I/O nodes: the inputs I and the output O such that I, O are subsets of N and I and O are disjoint.
- C is a set of constraints, called clocks.
- $fE : E \longrightarrow C$ is a mapping labeling each edge with a clock; it specifies the existence condition of the edges.

For instance, for the *counter* example:
$zv := v\$1 | v := (1 \text{ when } rst) \text{ default } zv + 1$
we get a SDD graph with:

- $N = \{1, v, zv + 1\}$
- $E = \{(1, v), (zv + 1, v)\}$
- $C = \{\widehat{rst}, \widehat{v} \wedge \neg\widehat{rst}\}$
- $fE((1, v)) = \widehat{rst}, fE((zv + 1, v)) = \widehat{v} \wedge \neg\widehat{rst}$

Let $SDD_1 =< G_1, C_1, fE_1 >$ and $SDD_2 =< G_2, C_2, fE_2 >$ to be two SDD graphs which represent the data dependency semantics of source and transformed programs, we say that SDD_2 is a *correct transformation* of SDD_1 on data dependency, or SDD_2*refines*SDD_1 w.r.t the data dependency semantics, if it satisfies that for any pair of nodes $x, y \in G_1 \cap G_2$ with $(x, y) \in E_1$:

- $fE_1(x, y)) \Rightarrow ((x, y) \in E_2 \wedge fE_2(x, y))$ (reinforcement)
- $(fE_2(x, y) \wedge fE_2(y, x) \Leftrightarrow \text{ false})$ (deadlock consistency)

6.6. Experiment with constraint-based testing

Participants: Christophe Junke, Loïc Besnard, Jean-Pierre Talpin.

Based on past experiences with contraint-based testing of Lustre programs, we investigated automatic test *sequences* generation for Signal: from a given test objective expressed as a boolean flow (or an event), we try to generate a sequence of inputs over discrete time which lead to an observation of the test objective. Our approach was based on an existing tool named GATeL, from CEA LIST, with the kind permission of its authors. This tool targets the Lustre language, so we reused Polychrony's Lustre generator to export Signal programs as Lustre nodes and use the result with GATeL to generate test sequences. The resulting test sequences were in turn formatted in a way suitable for simulation according to the original compilation of Signal to C: in other words, the generated sequences were tested on the actual program resulting from compilation of considered Signal specifications. During this experiment, we corrected Signal's Lustre generator tool which suffered from some several bugs that made it emit consistently incorrect Lustre programs. After some work, we could translate faithfully a little more than sixty existing Signal programs of simple to moderate complexity.

Our contribution is an example of how Signal can benefit from the pool of existing tools applicable to Lustre and why having a correct Signal-to-Lustre translator can be useful for Signal programs. This approach has its limits because it is not always possible nor adequate to fully translate a Signal program to Lustre: (1) By requiring the existence of one root clock and changing a program's input/output interface, it may be possible to simulate a Signal program in Lustre, but with loss of information (like user-defined flow dependencies); hence, some results based on the one Lustre implementation of a model may not easily be generalized to every possible execution of the original Signal program; (2) the complexity of Signal's semantics is mainly motivated by the power it gives to handle partial system specifications during the development process, whereas most Lustre tools expect fully defined executable programs; as such, they are of little help when dealing with most Signal programs. For those main reasons, it might be better to study and implement verification techniques around the Signal language and extend the set of formal tools that can reason about Signal programs.

More generally, our experiment can lead us to consider the use of constraint solving techniques with Signal, not only for verification but also compilation and simulation.

6.7. Polychronous modeling, analysis and validation for timed software architectures in AADL

Participants: Yue Ma, Huafeng Yu, Paul Le Guernic, Loïc Besnard, Thierry Gautier, Jean-Pierre Talpin.

High-level architecture modeling languages, such as AADL (Architecture Analysis and Design Language), are gradually adopted in the design of embedded systems so that design choice verification, architecture exploration and system property checking are carried out as early as possible. We are interested in the clock-based timing analysis, modeling and validation of software architectures specified in AADL [15]. In our approach, we first analyze the timing semantics of AADL, from which the formal polychronous/multiclock semantics is derived thanks to the multiclock nature of AADL specifications. Thus users are not suffered to find and/or build the fastest clock in the system. This distinguishes from [45], [37], where synchronous semantics is a prerequisite. This polychronous semantics is then expressed via a polychronous model of computation (MoC) [8] covering both AADL software, execution platform, and their binding. In addition, AADL thread-level scheduling is also explored and integrated according to affine clock relations [58]. In the framework of Polychrony, C or Java code is generated from the polychronous MoC. Simulation can then be carried out for the purpose of performance evaluation and verification.

Polychrony provides the back-end semantic-preserving transformation, scheduling, code generation, formal analysis and verification, architecture exploitation, and distribution [2]. With the scheduler synthesis, the translated AADL model is complete and executable, and can be used for the following analysis and validation [15]: 1) static analysis, including determinism identification and deadlock detection; 2) profiling-based analysis of real-time characteristics of a system [47]; 3) affine clock calculus to analyze the affine relations between clocks [58]; 4) co-simulation of AADL specifications and demonstration using the VCD technique [60]; 5) real-time scheduling and software/hardware allocation through the SynDEx tool [43].

An automatic toolchain dedicated to AADL modeling, scheduling, time analysis, verification, and simulation has been implemented and also integrated as plug-ins in the Eclipse framework. This toolchain (referred to as ASME2SSME) has been migrated from AADL V1.5.8 to AADL V2.0, together with OSATE V2. An experiment of interpretation of AADL Behavior Annex (BA) is initially performed, so that the Behavior Annex plugin is integrated in the modeling and transformation.

The whole model transformation and simulation chain has been migrated to Eclipse Indigo and attached to Polarsys as an Eclipse RCP. A tutorial case study, developed in the framework of the OPEES project [21], is adopted to illustrate the effectiveness of our contribution.

6.8. Static affine clocked-based scheduling and its seamless integration to ASME2SSME

Participants: Huafeng Yu, Yue Ma, Loïc Besnard, Thierry Gautier, Paul Le Guernic, Jean-Pierre Talpin.

An AADL model is not complete and executable if the thread-level scheduling is not resolved. Some scheduling tools, such as Cheddar [57], are well connected to AADL for schedulability analysis, scheduler synthesis and simulation inside these tools. However, they do not completely satisfy our demands for the following reasons: 1) logical and chronometric clocks are easily transformed into each other for formal and real-time analysis; 2) more events, such as input/output frozen events are also involved in the analysis; 3) static and periodic scheduling rather than stochastic/dynamic scheduling is expected due to predictability and formal verification; 4) the scheduling is easily and seamlessly connected to affine clock systems [58] so that formal analysis can be performed in Polychrony. Affine clock relations yield an expressive calculus for the specification and the analysis of time-triggered systems. A particular case of affine relations is the case of affine sampling relation expressed as $y = \{d \cdot t + \phi \mid t \in x\}$ of a reference discrete time x (d, t, ϕ are integers): y is a subsampling of positive phase ϕ and strictly positive period d on x.

We therefore propose a static affine-clocked-based scheduler synthesis process in the transformation from AADL to Signal, which includes the following subprocesses: 1) *calculate hyper-period* from the periods of all the threads according to the least common multiple principle; 2) *perform the scheduling* based on the hyper-period, and valid schedules are calculated according to a static, non-preemptive, and single-processor scheduling policy. More precisely, discrete events of each thread, such as dispatch, input/output frozen time, start and complete, are allocated in the hyper-period on condition that all their timing properties are satisfied. Affine clock relations of these events are ensured during the calculation. In the calculation process, different scheduling policies are considered, such as EDF and RM; 3) *export schedules to Signal affine clocks in a direct way*. This process, implemented as an independent Eclipse plugin, has been seamlessly integrated into the ASME2SSME toolchain.

6.9. Code distribution and architecture exploration via Polychrony and SynDEx

Participants: Huafeng Yu, Yue Ma, Loïc Besnard, Thierry Gautier, Jean-Pierre Talpin, Paul Le Guernic.

We propose an approach to address code distribution and multi-processor architecture exploration via the Polychrony and SynDEx toolchains. We consider high-level AADL models for the specification of multi-processor architecture. This architecture generally has a multiclock nature, thus it is modeled with a polychronous MoC. In this way, users are not suffered to find and/or build the fastest clock for a multi-processor architecture. According to this principle, AADL models are transformed into Signal models. To bridge between the polychronous semantics of Signal and synchronous semantics of SynDEx, clock synthesis in Polychrony [24] is applied. The translation from Signal to SynDEx is integrated in Polychrony. Finally, SynDEx models are used to perform distribution, scheduling, and eventually executive code generation for a specific architecture.

The main advantages of our approach are: 1) a formal model is adopted to connect the three languages, and it helps to preserve the semantic coherence and correct code generation in the transformations; 2) the formal model and methods used in the transformation are transparent to AADL practitioners, and it is fast and efficient to have the illustrative results for architecture exploration; 3) it provides the possibility for one of the three languages to take advantage of the functionalities provided by the other two languages. A toolchain has been developed, which includes model transformations between the three languages, considering both semantic and syntactic aspects. A tutorial case study, developed in the framework of the CESAR project [20], was adopted to demonstrate our contribution.

6.10. Design of safety-critical Java applications using affine abstract clocks

Participants: Adnan Bouakaz, Jean-Pierre Talpin.

Safety-critical Java (SCJ) is a domain specific API of Java that aims at the development of qualified and certified embedded systems. Despite its simplified memory and concurrency models, it is always difficult to ensure functional determinism and schedule feasibility when using shared-memory and traditional lock-based mutual exclusion protocols. Automated code generation techniques from data-flow specifications allow waiving part of the difficult and error-prone tasks of programming real-time schedules for computations and communications from the engineering process. Our ADFG tool aims at automatic SCJ code generation from data-flow specifications for a multitask implementation with an earliest-deadline first scheduler. The tool integrates the necessary formal analyses, model transformations, and the SCJ annotation checker as well.

The underlying data-flow model, called the affine data-flow (ADF) model of computation [14], is similar to cyclo-static data-flow graphs; it has however ultimately periodic production and consumption rates and a time-triggered operational semantics. We have also proposed a scheduling analysis of ADF specifications that consists of two major steps:

- The construction of abstract affine schedules for computations that minimize buffering requirements under the assumption of read-write precedences and exclude overflow and underflow exceptions over communication channels. Affine transformations of clocks were first introduced in the context of Signal programs [58] and used in the ADF model to relate the activation rates of connected actors.
- The concretization of the affine schedules using an earliest-deadline first (EDF) symbolic schedulability analysis in a way that read-write precedences is ensured without the need for lock-based mechanisms and the processor utilization factor is maximized.

6.11. Polychronous controller synthesis from MARTE CCSL timing specifications

Participants: Huafeng Yu, Loïc Besnard, Thierry Gautier, Jean-Pierre Talpin, Paul Le Guernic.

CCSL (Clock Constraint Specification Language) [29] is defined in an annex of the UML MARTE profile [53]. We are interested in the analysis, synthesis, code generation and simulation of polychronous systems specified in CCSL. Timed systems subject to clock expressions or relations can be modeled, specified, analyzed, and simulated within software environments such as SCADE [40], TimeSquare [44] and Polychrony. The motivation of our work, to address the simulation and code generation of polychronous systems, is to take advantage of the formal framework of Polychrony in the context of a high-level specification formalism, MARTE CCSL [62]. Yet, our work considers a novel approach with regard to previous approaches: to generate executable specifications by considering discrete controller synthesis (DCS) [56], [51], [52].

Based on our previous work on clock hierarchization [61] and the general clock synthesis approach [62], our current work concentrates on the study of interface-oriented clock synthesis in the context of distributed components. In this work, CCSL clock constraints are specified on the clocked signals that pass through the interface, and the controller to synthesize is used to ensure the constraints. Interface-level synthesis helps to reduce the synthesis complexity since communication concerns and internal component behavior are isolated from the synthesis. The controllability analysis of signals and clock relations are studied with regard to endochronous, polychronous, and reactive components. This analysis leads to the separation of controllable and uncontrollable signals in the synthesis. Observers of CCSL clock constraints have been proposed in order to specify control objectives. In addition, properties of local components and the global system, such as determinism and deadlock, are also initially studied.

6.12. An integration language for Averest/Quartz and Polychrony/Signal

Participants: Ke Sun, Jean-Pierre Talpin.

As typical synchronous languages, Quartz and Signal possess their own respective characteristics [11]. In particular, Quartz, an imperative synchronous language, mainly focuses on the description of the control-flow. The Quartz model is always reactive and synchronously deterministic. Different from Quartz models which can only be monochronous, a process in Signal may be polychronous, meaning that its clock hierarchy can form a forest. Therefore, the potential integration between Averest, a framework for Quartz, and Polychrony, a toolset for Signal, may offer a practical mode to develop globally asynchronous locally synchronous (GALS) systems: program imperative and reactive modules in Quartz, then synthesize the scheduler from their Signal network specification.

To maximally benefit from the existing achievements for the two languages [12], the main idea is to communicate the Quartz modules with each other via asynchronous wrappers without changing the original code. Considering that the Quartz modules should be still deterministic in asynchronous environment, the wrapper should be capable of controlling the IO streams. On the other hand, the wrapper, as a module interface, will make sense for automatic scheduler synthesis, the next step.

We will propose a new, easy to use, domain-specific language to help the user specify the input traces as requirements to the environment and define the IO traces as guarantee of the module. From the user-defined specification, a series of clock constraints, assertions, etc. may be synthesized in the form of Signal specification. Thus, this language may bridge the gap between Polychrony/Signal and Averest/Quartz.

FORMES Team

6. New Results

6.1. Higher-Order Abstract Syntax

This recently started project funded by the National Science Foundation of China aims at setting up a generic infrastructure for representating logical systems and automate their meta-theoretical study. We view a logical system as a type theory made of three components: a language of terms, types being particular terms; a set of typing rules; and a set of computational rules described by typed higher-order rewrite rules.

There are several challenges in this project. The first is to define logical frameworks which are expressive enough -at least as expressive as Girard's System F or Edingburgh's LF- to define the syntax and semantics of rich type theories, such as CoqMTU as an extreme example. A second challenge is to develop new techniques for checking the three main properties of higher-order rewrite rules: type preservation -which is usually easy-, confluence and termination. Our work here has progressed steadily, in paticular with new advanced techniques for checking termination and confluence described next. A third challenge is to formalize these results in Coq, in order to provide proof certificates for particular cases. The fourth challenge is to build a a general infrastructure in Coq in which all these techniques become available in order to study particular logical systems.

As initial steps, we undertook the following formalizations :

- Hua Mei implemented an intensional framework for simply typed lambda-calculus in Coq, where α- and β-conversions have been axiomized.
- Frédéric Blanqui has formalized in Coq the pure lambda-calculus following the definition of Curry and Feys in [43] (named variables and explicit alpha-equivalence), and the proof of termination of β-reduction for simply-typed λ-terms based on computability predicates [51]. To the best of his knowledge, this is the first formalization of the termination of β-reduction using named variables and explicit alpha-equivalence, all the other formalizations using De Bruijn indices [73] or nominal logic [48].
- Qian Wang formalized completely the theory of CoqMTU in Coq augmented with strong set-theoretic axioms in order to get around Gödel's incompleteness theorem. This is described in more details next.

6.2. CoqMTU

The proof-assistant Coq is based on a complex type theory, which resulted from various extensions of the Calculus of Constructions studied independently fromf each other. With Bruno Barras, we decided to address the challenge of proving the real type theory underlying Coq, and even, indeed, its recent extension CoqMT. To this end, we have studied formally the theory CoqMTU, which extends the calculus of Constructions with inductive types, a predicative hierarchy of universes and a decidable theory T for some first-order inductive types for which large elimination is no more available. This work has been published at LICS [1]. It leaves open the question whether large elimination can be accomadated for those inductive types which carry along a decidable theory T. This problem has been solved recently by Wang, who constructed a set-theoretic model of CoqMTU with strong elimination.

6.3. Normal Rewriting

There are many forms of rewriting used in the litterature: plain rewriting (rules are fired via plain pattern matching), rewriting modulo T (rules are fired via pattern matching modulo T), higher-order rewriting (rules are fired via higher-order pattern matching, but apply to simply typed lambda-terms terms provided the redex is of base type and in beta-normal eta-long form). For each of these rewriting mechanisms, there are results describing how to check confluence and termination.

Regarding confluence, these results describe which *critical pairs* must be computed in order to check the confluence property of the rewriting relation, assuming some termination property. In [17], we describe a general abstract result which can then be instantiated to all of the previous cases, and removes the assumptions above for higher-order rewriting. This is done via two novel notions: abstract positional rewriting allows us to capture the notion of critical peak without having to talk about a specific term structure; abstract normal rewriting with a triple (R, S, E) allows us to capture all different forms of rewriting: $S = E = \varnothing$ for plain rewriting; $S = \varnothing$ for rewriting modulo; E is alpha-conversion for higher-order rewriting, while the set of simplifiers S is made of beta-reduction and eta-expansion, R being the set of user-defined rules. Of course, there are other applications of normal rewriting described in the paper: for first-order computations, but also for higher-order computations at higher types, or using eta-reduction instead of eta-expansion, therefore solving a long-standing open problem.

Regarding termination, these results are very preliminary. In a recent paper submitted to ACM Transactions on Computational Logics, we extend the termination proof methods for higher-order computations based on plain pattern matching to higher-order rewriting systems based on higher-order pattern matching. We accomodate, for the one hand, with a weakly polymorphic, algebraic extension of Church's simply typed λ-calculus, and on the other hand, with any use of eta, as a reduction, as an expansion or as an equation. User's rules may be of any type in this type system, either a base, functional, or polymorphic type. Our techniques fit well with higher-order reduction orderings, such as the computability path ordering, but can also be used by other techniques, such as higher-order dependency pairs. All examples of normal higher-order rewrite rules that can be found in the litterature can be treated by our techniques, even those for which termination is by no means obvious to the expert.

6.4. Decreasing Diagrams

Based on the so-called Newman's lemma, the method for checking confluence introduced in the former paragraph applies to terminating computations. A completely different technique based on the so-called Hindley-Rosen's lemma applies when computation do not terminate, and is at the basis of Tait's confluence proof for the pure lambda-calculus. In recent papers, van Oostrom succeeded to capture both within a single framework thanks to the notion of decreasing diagram of a labelled abstract relation [76], see also [11] for an improved proof. Decreasing diagrams are specific convertibility proofs for local peaks, which labels are smaller in some sense than those of the local peak they aim at replacing. Any convertibility proof can then be converted into a confluence proof by recursively replacing its local peaks by their associated decreasing diagrams. Using a subtle characterization of confluence for arbitrary (possibly non-terminating) relations by cofinal derivations due to Klop [11], van Oostrom showed that any confluent relation which convertibility classes are countable, can be labelled in a way that makes it a labelled relation satisfying the decreasing diagram condition.

In [15], we first give a new, simple proof of van Oostrom's initial result based on a subtle well-founded order on conversions, and generalize it to rewriting modulo by using *strongly coherent cliffs* as an analog of decreasing diagrams for peaks. We then extend Klop's cofinal derivations to *cofinal streams*, and prove again a completeness result under the strong coherence assumption. Finally, we derive from these results a new, compact proof of Toyama's theorem that confluence is a modular property of rewriting systems built on disjoint vocabularies, and extend it to rewriting modulo when strong coherence is satisfied.

We are now trying to get rid of the strong coherence assumption by introducing a weaker analog of decreasing diagrams, *decreasing cliffs*. A preliminary result was presented early november at the Japanese Term Rewriting Workshop in Sendai.

This line of work is very promising. We expect it will eventually lead to the solution of an old open problem, the characterization of a class of non-left linear, non-terminating rewrite systems for which confluence is decidable by means of (parallel) critical pairs. We believe that the implementation of such a result would be impact the way confluence proofs are carried out, including in type theory.

6.5. Higher-order Reduction Orderings

Since HORPO , several higher-order reduction orderings have been described, based on either Dershowitz's RPO , Blanqui-Jouannaud-Okada's Computational Closure , and Arts and Giesel' dependency pairs . Our work continues in three different directions:

- CPO is an order for simply typed lambda-terms that allows to show strong normalization of beta-reduction even in presence of higher-order rewrite rules provided these rules decrease in the ordering [32]. It is currently the only automated mechanism that achieves non-trivial computations by turning Girard's computability predicates method into a usable tool. It has been shown that CPO can handle weakly polymorphic type disciplines, as well as inductive types. Recently, we have shown that CPO scales up to dependently typed calculi as LF. We are currently writing a paper describing CPO and its extensions to calculi with inductive and dependent types which should be submitted to a journal by the end of the year.
- Frédéric Blanqui defended his "Habilitation à diriger des recherches" at the University Denis Diderot (Paris 7) on July 13. In [13], he gives a synthetic view on how the notion of computability closure can be used to prove the termination of various kinds of rewrite relations (class rewriting or rewriting with matching modulo), and how it relates with other notions (dependency pairs, semantic labeling, and HORPO, the predecessor of CPO.
- Frédéric Blanqui has developed an automated termination prover called HOT based on the above work on the computability closure and his former work on size annotations [31]. For its first participation, HOT won the international competition on termination in the category "higher-order rewriting union beta".

6.6. Certification of Termination Proofs

Frédéric Blanqui and Kim Quyen Ly continued to work on the development of a new version of Rainbow based on Coq extraction mechanism [59]. We developed a tool generating from an XSD file, Coq and OCaml data structures representing the XML types defined the XSD file, and OCaml parsing functions for generating such data structures from an XML file. The main difficulty was to topologically reorder the XSD type definitions in order to get simple and well defined Coq data structures. We also defined and proved in Coq a function for checking the correctness of termination certificates based on the DP transformation [26]. The main difficulty was to manage the evolution of the arity function along the transformation. Indeed, to simplify the translation of CPF elements into the data structures used in CoLoR [30], we decided to use a fixed but infinite set of symbols [69]. However the arity function need to be updated along the transformations applied to the system. These results are presented in [20].

6.7. Certification of Moca

Frédéric Blanqui has formalized in Coq and proved the correctness and completeness of the construction functions generated by Moca for the theory of groups [29]. The first difficulty is to represent the Moca functions themselves in a faithful way because, in Coq, there is no "when" clauses and "match" constructions are expanded into elementary "case" constructions with no tuple patterns and patterns of depth one only. In addition, Coq termination checker only accepts functions with exactly one structurally decreasing argument, which is generally not the case of Moca functions. The second difficulty is the completeness proof: it requires the use of intermediate data structures for reasoning on normal forms. During his internship, Rémi Nollet (L3, ENS Lyon) improved the representation of OCaml functions by using inductive predicates, and extended the correctness proof to commutative groups.

6.8. First steps towards the certification of an ARM simulator

The simulation of Systems-on-Chip (SoC) is nowadays a hot topic because, beyond providing many debugging facilities, it allows the development of dedicated software before the hardware is available. Low-consumption

CPUs such as ARM play a central role in SoC. However, the effectiveness of simulation depends on the faithfulness of the simulator. To this effect, we started to prove significant parts of such a simulator, SimSoC. Basically, on one hand, we develop a Coq formal model of the ARM architecture while on the other hand, we consider a version of the simulator including components written in Compcert-C [58]. Then we prove that the simulation of ARM operations, according to Compcert-C formal semantics, conforms to the expected formal model of ARM. Size issues are partly dealt with using automatic generation of significant parts of the Coq model and of SimSoC from the official textual definition of ARM [3]. A second step was achieved in [12], with the proof a significant instruction (ADC, Add with Carry). A crucial technical issue was then raised: facilitating reasoning by inversion on the rules defined in Compcert-C. Hundreds such steps are required for a single instruction, and each of them generates a dozen of new names. Relying on Coq tactic inversion results in unmanageable scripts, very fragile and difficult to maintain. In 2012 we dealt with this issue by designing our own inversion mechanism, allowing us to improve automation of the proof, while keeping enough command so that interactive steps refer to controlled names. It was then possible to get a much shorter proof on ADC and to prove at least one instruction in each category of the ARM instruction set.

6.9. Certified implementation of BIP

BIP (*Behavior, Interaction, Priority*) is a component-based language designed at VERIMAG for modeling and programming complex embedded systems [27]. A BIP model is essentially a set of atomic components described with explicit states and transitions, composed together in a hierarchical way. The main original feature of BIP lies in a very rich notion of *connector* for defining interactions between components [33]. An efficient implementation of BIP in C++ is already available at VERIMAG.

Building on our previous experience on SimSoC, we started to work on a certified implementation of BIP. Our long term objective is to propose a certified compilation chain from BIP models to embedded code, through a first translation from BIP to Compcert-C.

In 2012 we focused on a simple subset of BIP Currently, we have a first definition of a formal semantics of this subset in Coq, in two versions: an relational version, inspired by a rule-based operational semantics, and a functional version, which specifies a possible implementation of the relational version (in particular, it includes a scheduler). We also produce a Compcert-C code which is expected to behave exactly like the functional semantics, and we started to state and prove corresponding statements on very simple BIP models.

6.10. Formal model and proofs for Netlog protocols

Netlog is a language designed and implemented in the Netquest project for describing protocols. Netlog has a precise semantics, provides a high level of abstraction thanks to its Datalog flavor and benefits from an efficient implementation. This makes it a very interesting target language for proofs of protocols.

Jean-François Monin, Stéphane Grumbach (formerly LIAMA/Netquest) and Yuxin Deng (Jiaotong University, Shanghai) designed a formal model of Netlog in Coq, where the two possible semantics are derived from common basic blocks. In a fully certified framework, a formal proof of the Netlog engine (running on each node) would be required. We don't attack this part at the moment: we assume that the implementation respects the general properties stated in our model and focus on the issues raised by the distributed model of computation provided by Netlog. This framework could be applied to an algorithm constructing a Breadth-First Search Spanning Tree (BFS) in a distributed system [45].

In 2011, Jean-François Monin and Meixian Chen (Jiaotong Shanghai) generalized the model in order to take the removal of datalog facts into account, and used the improved framework to Prim's algorithm. In 2012, this work was slightly improved and published in [16].

6.11. Formalisation of security APIs for mobile phones

This work is in cooperation with Nokia Beijing, who was interested by the application of verification technologies to mobile phones. We decided to focus on security APIs, considering that mobile devices are commonly used by end-users to store their personnal data (e.g., passwords), while running all sort of downloaded applications at the same time.

For 2012, we (including Nokia) agreed to consider devices under Android, though Nokia switched to windows, in order to circumvent copyright issues.

Three models and corresponding sets of APIs for password storage applications on Android were developed. Each model fixes some bugs of the previous one and introduces a new feature. We consider the third model is enough for the basic function and well built to be safe. Then, a full Coq proof of the third model was developed as well as its corresponding API's security property. A suitable abstraction of the application on the phone within its environment is described as a state transition system. Then we proved by induction that the expected secrets actually remain secret at any reachable state.

6.12. Trace Analysis

Simulation sessions produce huge trace files, sometimes now in hundreds of gigabytes, that are hard to analyze with a quick response time. This comes down to two sub-problems:

- The trace file size. Trace files are huge because they include lots of information. But when looking for a specific problem, one does not need all of this information. To search one given defect, one may ignore a large amount of the data in the trace file. One would like the trace file to contain only relevant information to the concerned problem.
- The expressive power of the language to analyze the trace, and its usability. If the language is limited to expression search, it is easy to use but hard to construct sophisticated formulas. If the language used is Linear Temporal Logic (LTL), there is a very high expressive power but many engineers are unable to write a LTL formula and to maintain it over time.

We have started to build a trace analysis tool. It includes a language which allows expression of time-related formulas as a subset of LTL, but is simple to formulate expressions. When this language is compiled, the compiler generates two outputs:

- a filter script that will help reduce the size of the trace file.
- a program that analyzes such trace files to find whether the formula is satisfied.

When compiling one trace language input file, it generates a filter script. The filter script is a set of data descriptors. It describes which events from the simulator must be traced and which should be ignored. Then during the simulation, the filter is loaded and only the required output is generated.

We have started to design a trace language and a compiler, and extended the SimSoC simulator to support generation of trace files with a filter. A first version of the trace language compiler has been implemented in OCAML, which generates OCAML programs for trace analysis. In the current version under development, the filters are not yet parallelized with simulation.

GALAAD Project-Team

6. New Results

6.1. Algebraic representations for geometric modeling

6.1.1. Fitting ideals and multiple-points of surface parameterizations

Participants: Nicolàs Botbol, Laurent Busé.

Given a birational parameterization ϕ of an algebraic surface $\mathcal{S}$ in the projective space $\mathbb{P}^3$, the purpose of this ongoing work is to investigate the sets of points $D_k(\phi)$ on $\mathcal{S}$ whose preimage consists in k or more points, counting multiplicity. Our main result is an explicit description of these algebraic sets $D_k(\phi)$ in terms of Fitting ideals of some graded parts of a symmetric algebra associated to the parameterization ϕ.

This work is done in collaboration with Marc Chardin (University Pierre et Marie Curie).

6.1.2. Algebraic geometry tools for the study of entanglement: an application to spin squeezed states

Participant: Alessandra Bernardi.

In [18] a short review of Algebraic Geometry tools for the decomposition of tensors and polynomials is given from the point of view of applications to quantum and atomic physics. Examples of application to assemblies of indistinguishable two-level bosonic atoms are discussed using modern formulations of the classical Sylvester's algorithm for the decomposition of homogeneous polynomials in two variables. In particular, the symmetric rank and symmetric border rank of spin squeezed states is calculated as well as their Schrödinger-cat-like decomposition as the sum of macroscopically different coherent spin states; Fock states provide an example of states for which the symmetric rank and the symmetric border rank are different.

This is a joint work with I. Carusotto (University of Trento, Italy).

6.1.3. A partial stratification of secant varieties of Veronese varieties via curvilinear subschemes.

Participant: Alessandra Bernardi.

In [11] we give a partial quasi-stratification of the secant varieties of the order d Veronese variety $X_{m,d}$ of $\mathbb{P}^m$. It covers the set $\sigma_t(X_{m,d})^\dagger$ of all points lying on the linear span of curvilinear subschemes of $X_{m,d}$, but two quasi-strata may overlap. For low border rank, two different quasi-strata are disjoint and we compute the symmetric rank of their elements. Our tool is the Hilbert schemes of curvilinear subschemes of Veronese varieties. To get a stratification we attach to each $P \in \sigma_t(X_{m,d})^\dagger$ the minimal label of a quasi-stratum containing it.

This is a joint work with E. Ballico (University of Trento, Italy).

6.1.4. Decomposition of homogeneous polynomials with low rank.

Participant: Alessandra Bernardi.

Let F be a homogeneous polynomial of degree d in $m+1$ variables defined over an algebraically closed field of characteristic 0 and suppose that F belongs to the s-th secant variety of the d-uple Veronese embedding of $\mathbb{P}^m$ into $\mathbb{P}^{\binom{m+d}{d}-1}$ but that its minimal decomposition as a sum of d-th powers of linear forms $M_1, ..., M_r$ is $F = M_1^d + \cdots + M_r^d$ with $r > s$. In [12], we show that if $s + r \leq 2d + 1$ then such a decomposition of F can be split into two parts: one of them is made by linear forms that can be written using only two variables. The other part is uniquely determined once one has fixed the first part. We also obtain a uniqueness theorem for the minimal decomposition of F if r is at most d and a mild condition is satisfied.

This is a joint work with E. Ballico (University of Trento, Italy).

6.1.5. Higher secant varieties of $\mathbb{P}^n \times \mathbb{P}^1$ embedded in bi-degree (a, b)

Participant: Alessandra Bernardi.

In [15], we compute the dimension of all the higher secant varieties to the Segre-Veronese embedding of $\mathbb{P}^n \times \mathbb{P}^1$ via the section of the sheaf $\mathcal{O}(a, b)$ for any $n, a, b \in \mathbb{Z}^+$. We relate this result to the Grassmann Defectivity of Veronese varieties and we classify all the Grassmann $(1, s-1)$-defective Veronese varieties.

This is a joint work with E. Ballico, M. V. Catalisano (University of Trento, Italy).

6.1.6. Symmetric tensor rank with a tangent vector: a generic uniqueness theorem

Participant: Alessandra Bernardi.

Let $X_{m,d} \subset \mathbb{P}^N$, $N := \binom{m+d}{m} - 1$, be the order d Veronese embedding of $\mathbb{P}^m$. Let $\tau(X_{m,d}) \subset \mathbb{P}^N$, be the tangent developable of $X_{m,d}$. For each integer $t \geq 2$ let $\tau(X_{m,d}, t) \subseteq \mathbb{P}^N$, be the join of $\tau(X_{m,d})$ and $t-2$ copies of $X_{m,d}$. In [13], we prove that if $m \geq 2$, $d \geq 7$ and $t \leq 1 + \lfloor \binom{m+d-2}{m} / (m+1) \rfloor$, then for a general $P \in \tau(X_{m,d}, t)$ there are uniquely determined $P_1, \cdots, P_{t-2} \in X_{m,d}$ and a unique tangent vector ν of $X_{m,d}$ such that P is in the linear span of $\nu \cup \{P_1, \cdots, P_{t-2}\}$. In other words, a degree d linear form f (a symmetric tensor T of order d) associated to P may be written as

$$f = L_{t-1}^{d-1} L_t + \sum_{i=1}^{t-2} L_i^d, \quad (T = v_{t-1}^{\otimes(d-1)} v_t + \sum_{i=1}^{t-2} v_i^{\otimes d})$$

with L_i linear forms on $\mathbb{P}^m$ (v_i vectors over a vector field of dimension $m+1$ respectively), $1 \leq i \leq t$, that are uniquely determined (up to a constant).

This is a joint work with E. Ballico (University of Trento, Italy).

6.1.7. General tensor decomposition, moment matrices and applications.

Participants: Alessandra Bernardi, Bernard Mourrain.

In [17] the tensor decomposition addressed may be seen as a generalisation of Singular Value Decomposition of matrices. We consider general multilinear and multihomogeneous tensors. We show how to reduce the problem to a truncated moment matrix problem and give a new criterion for flat extension of Quasi-Hankel matrices. We connect this criterion to the commutation characterisation of border bases. A new algorithm is described. It applies for general multihomogeneous tensors, extending the approach of J.J. Sylvester to binary forms. An example illustrates the algebraic operations involved in this approach and how the decomposition can be recovered from eigenvector computation.

This is a joint work with J. Brachat and P. Comon (i3S, CNRS).

6.1.8. On the cactus rank of cubic forms

Participant: Alessandra Bernardi.

In this work, we prove that the smallest degree of an apolar 0-dimensional scheme of a general cubic form in $n+1$ variables is at most $2n+2$, when $n \geq 8$, and therefore smaller than the rank of the form. For the general reducible cubic form the smallest degree of an apolar subscheme is $n+2$, while the rank is at least $2n$.

This is a joint work with K. Ranestad (University of Oslo, Norway) that will be published in 2013 in the Journal of Symbolic Computation. The preprint is available at http://hal.inria.fr/inria-00630456.

6.1.9. Tensor ranks on tangent developable of Segre varieties

Participant: Alessandra Bernardi.

In [14] we describe the stratification by tensor rank of the points belonging to the tangent developable of any Segre variety. We give algorithms to compute the rank and a decomposition of a tensor belonging to the secant variety of lines of any Segre variety. We prove Comon's conjecture on the rank of symmetric tensors for those tensors belonging to tangential varieties to Veronese varieties.

This is a joint work with E. Ballico (University of Trento, Italy).

6.1.10. On the dimension of spline spaces on planar T-meshes

Participant: Bernard Mourrain.

In [33], we analyze the space of bivariate functions that are piecewise polynomial of bi-degree $\leq (m, m')$ and of smoothness r along the interior edges of a planar T-mesh. We give new combinatorial lower and upper bounds for the dimension of this space by exploiting homological techniques. We relate this dimension to the weight of the maximal interior segments of the T-mesh, defined for an ordering of these maximal interior segments. We show that the lower and upper bounds coincide, for high enough degrees or for hierarchical T-meshes which are enough regular. We give a rule of subdivision to construct hierarchical T-meshes for which these lower and upper bounds coincide. Finally, we illustrate these results by analyzing spline spaces of small degrees and smoothness.

6.1.11. On the problem of instability in the dimension of a spline space over a T-mesh

Participant: Bernard Mourrain.

In [23], we discuss the problem of instability in the dimension of a spline space over a T-mesh. For bivariate spline spaces $S(5,5,3,3)$ and $S(4,4,2,2)$, the instability in the dimension is shown over certain types of T-meshes. This result could be considered as an attempt to answer the question of how large the polynomial degree (m, m') should be relative to the smoothness (r, r') to make the dimension of a spline space stable. We show in particular that the bound $m \geq 2r+1$ and $m' \geq 2r'+1$ are optimal.

This is a joint work with Berdinsky Dmitry, Oh Min-Jae and Kim Taewan (Department of Naval Architecture and Ocean Engineering Seoul National University, South Korea).

6.1.12. Homological techniques for the analysis of the dimension of triangular spline spaces

Participant: Bernard Mourrain.

The spline space $C_k^r(\Delta)$ attached to a subdivided domain Δ of $\mathbb{R}^d$ is the vector space of functions of class C^r which are polynomials of degree $\leq k$ on each piece of this subdivision. Classical splines on planar rectangular grids play an important role in Computer Aided Geometric Design, and spline spaces over arbitrary subdivisions of planar domains are now considered for isogeometric analysis applications. In [34], we address the problem of determining the dimension of the space of bivariate splines $C_k^r(\Delta)$ for a triangulated region Δ in the plane. Using the homological introduced by Billera (1988), we number the vertices and establish a formula for an upper bound on the dimension. There is no restriction on the ordering and we obtain more accurate approximations to the dimension than previous methods. Furthermore, in certain cases even an exact value can be found. The construction makes it also possible to get a short proof for the dimension formula when $k \geq 4r+1$, and the same method we use in this proof yields the dimension straightaway for many other cases.

This is a joint work with Nelly Villamizar (CMA, University of Oslo, Norway).

6.1.13. Analysis-suitable volume parameterization of multi-block computational domain in isogeometric applications

Participants: Bernard Mourrain, André Galligo.

Parameterization of computational domain is a key step in isogeometric analysis just as mesh generation is in finite element analysis. In [36], we study the volume parameterization problem of multi-block computational domain in isogeometric version, i.e., how to generate analysis-suitable parameterization of the multi-block computational domain bounded by B-spline surfaces. Firstly, we show how to find good volume parameterization of single-block computational domain by solving a constraint optimization problem, in which the constraint condition is the injectivity sufficient conditions of B-spline volume parametrization, and the optimization term is the minimization of quadratic energy functions related to the first and second derivatives of B-spline volume parameterization. By using this method, the resulted volume parameterization has no self-intersections, and the isoparametric structure has good uniformity and orthogonality. Then we extend this method to the multi-block case, in which the continuity condition between the neighbor B-spline volume should be added to the constraint term. The effectiveness of the proposed method is illustrated by several examples based on three-dimensional heat conduction problem.

This is a joint work with Régis Duvigneau (Inria, EPI OPALE) and Xu Gang (College of computer - Hangzhou Dianzi University, China).

6.1.14. A new error assessment method in isogeometric analysis of 2D heat conduction problems

Participants: Bernard Mourrain, André Galligo.

In [35], we propose a new error assessment method for isogeometric analysis of 2D heat conduction problems. A posteriori error estimation is obtained by resolving the isogeometric analysis problem with several k-refinement steps. The main feature of the proposed method is that the resulted error estimation surface has a B-spline form, according to the main idea of isogeometric analysis. Though the error estimation method is expensive, it can be used as an error assessment method for isogeometric analysis. Two comparison examples are presented to show the efficiency of the proposed method.

This is a joint work with Régis Duvigneau (Inria, EPI OPALE) and Xu Gang (College of computer - Hangzhou Dianzi University, China).

6.1.15. On the cut-off phenomenon for the transitivity of randomly generated subgroups

Participant: André Galligo.

Consider $K \geq 2$ independent copies of the random walk on the symmetric group S_N starting from the identity and generated by the products of either independent uniform transpositions or independent uniform neighbor transpositions. At any time $n \in \mathbb{N}$, let G_n be the subgroup of S_N generated by the K positions of the chains. In the uniform transposition model, we prove in [28] that there is a cut-off phenomenon at time $N\ln(N)/(2K)$ for the non-existence of fixed point of G_n and for the transitivity of G_n, thus showing that these properties occur before the chains have reached equilibrium. In the uniform neighbor transposition model, a transition for the non-existence of a fixed point of G_n appears at time of order $N^{1+\frac{2}{K}}$ (at least for $K \geq 3$), but there is no cut-off phenomenon. In the latter model, we recover a cut-off phenomenon for the non-existence of a fixed point at a time proportional to N by allowing the number K to be proportional to $\ln(N)$. The main tools of the proofs are spectral analysis and coupling techniques.

This is a joint work with Laurent Miclo (University of Toulouse).

6.2. Algebraic algorithms for geometric computing

6.2.1. On the isotopic meshing of an algebraic implicit surface

Participant: Bernard Mourrain.

In [22], we present a new and complete algorithm for computing the topology of an algebraic surface given by a squarefree polynomial in $\mathbb{Q}[X, Y, Z]$. Our algorithm involves only subresultant computations and entirely relies on rational manipulation, which makes it direct to implement. We extend the work in [15], on the topology of non-reduced algebraic space curves, and apply it to the polar curve or apparent contour of the surface S. We exploit simple algebraic criterion to certify the pseudo-genericity and genericity position of the surface. This gives us rational parametrizations of the components of the polar curve, which are used to lift the topology of the projection of the polar curve. We deduce the connection of the two-dimensional components above the cell defined by the projection of the polar curve. A complexity analysis of the algorithm is provided leading to a bound in $\widetilde{\mathcal{O}}_B(d^{15}\tau)$ for the complexity of the computation of the topology of an implicit algebraic surface defined by integer coefficients polynomial of degree d and coefficients size τ . Examples illustrate the implementation in Mathemagix of this first complete code for certified topology of algebraic surfaces.

This is a joint work with Daouda Niang Diatta, Olivier Ruatta (XLIM, University of Limoges).

6.2.2. Moment matrices, border basis and real radical computation

Participant: Bernard Mourrain.

In [32], we describe new methods to compute the radical (resp. real radical) of an ideal, assuming its complex (resp. real) variety is finite. The aim is to combine approaches for solving a system of polynomial equations with dual methods which involve moment matrices and semi-definite programming. While border basis algorithms are efficient and numerically stable for computing complex roots, algorithms based on moment matrices allow the incorporation of additional polynomials, e.g., to restrict the computation to real roots or to eliminate multiple solutions. The proposed algorithm can be used to compute a border basis of the input ideal and, as opposed to other approaches, it can also compute the quotient structure of the (real) radical ideal directly, i.e., without prior algebraic techniques such as Gröbner bases. It thus combines the strengths of existing algorithms and provides a unified treatment for the computation of border bases for the ideal, the radical ideal and the real radical ideal.

This is a joint work with Jean-Bernard Lasserre (LAAS, Toulouse), Monique Laurent (CWI, Amsterdam, Netherland), Philipp Rostalski (University of California, Berkeley, US) Philippe Trébuchet (APR, LIP6, Paris).

6.2.3. On the computation of matrices of traces and radicals of ideals

Participant: Bernard Mourrain.

Let $f_1, ..., f_s \in \mathbb{K}[x_1, ..., x_m]$ be a system of polynomials generating a zero-dimensional ideal I, where $\mathbb{K}$ is an arbitrary algebraically closed field. In [31], we study the computation of “matrices of traces” for the factor algebra $\mathcal{A} := \mathbb{C}[x_1, ..., x_m]/I$, i.e. matrices with entries which are trace functions of the roots of I. Such matrices of traces in turn allow us to compute a system of multiplication matrices $\{M_{x_i} | i = 1, ..., m\}$ of the radical $\sqrt{I}$. We first propose a method using Macaulay type resultant matrices of $f_1, ..., f_s$ and a polynomial J to compute moment matrices, and in particular matrices of traces for $\mathcal{A}$. Here J is a polynomial generalizing the Jacobian. We prove bounds on the degrees needed for the Macaulay matrix in the case when I has finitely many projective roots in $\mathbb{P}^m_{\mathbb{C}}$. We also extend previous results which work only for the case where $\mathcal{A}$ is Gorenstein to the non-Gorenstein case. The second proposed method uses Bezoutian matrices to compute matrices of traces of $\mathcal{A}$. Here we need the assumption that $s = m$ and $f_1, ..., f_m$ define an affine complete intersection. This second method also works if we have higher dimensional components at infinity. A new explicit description of the generators of $\sqrt{I}$ are given in terms of Bezoutians.

This is a joint work with Itnuit Janovitz-Freireich (Departamento de Matemáticas, Mexico), Lajos Ronayi (Hungarian Academy of Sciences and Budapest University of Technology and Economics, Budapest), Agnes Szanto (Department of Computer Science, North Carolina State University, US).

6.2.4. Border basis representation of a general quotient algebra

Participant: Bernard Mourrain.

In [40], we generalized the construction of border bases to non-zero dimensional ideals for normal forms compatible with the degree, tackling the remaining obstacle for a general application of border basis methods. First, we give conditions to have a border basis up to a given degree. Next, we describe a new stopping criterion to determine when the reduction with respect to the leading terms is a normal form. This test based on the persistence and regularity theorems of Gotzmann yields a new algorithm for computing a border basis of any ideal, which proceeds incrementally degree by degree until its regularity. We detail it, prove its correctness, present its implementation and report some experimentations which illustrate its practical good behavior.

This is a joint work with Philippe Trébuchet (APR, LIP6, Paris).

6.2.5. Voronoï diagrams of algebraic distance fields

Participant: Bernard Mourrain.

In [25], we design and implement an efficient and certified algorithm for the computation of Voronoi Diagrams (VD's) constrained to a given domain. Our framework is general and applicable to any VD-type where the distance field is given explicitly or implicitly by a polynomial, notably the anisotropic VD or VD's of non-punctual sites. We use the Bernstein form of polynomials and DeCasteljau's algorithm to subdivide the initial domain and isolate bisector, or domains that contain a Voronoi vertex. The efficiency of our algorithm is due to a filtering process, based on bounding the field over the subdivided domains. This allows us to exclude functions (thus sites) that do not contribute locally to the lower envelope of the lifted diagram. The output is a polygonal description of each Voronoi cell, within any user-defined precision, isotopic to the exact VD. Correctness of the result is implied by the certified approximations of bisector branches, which are computed by existing methods for handling algebraic curves. First experiments with our C++ implementation, based on double precision arithmetic, demonstrate the adaptability of the algorithm.

This is a joint work with Ioannis Emiris (ERGA, National Kapodistrian University of Athens, Greece), Angelos Mantzaflaris (RICAM, Austrian Academy of Sciences, Austria).

6.2.6. Rational invariants of scalings from Hermite normal forms

Participant: Evelyne Hubert.

Scalings form a class of group actions that have both theoretical and practical importance. A scaling is accurately described by an integer matrix. In [39] tools from linear algebra are exploited to compute a minimal generating set of rational invariants, trivial rewriting and rational sections for such a group action. The primary tools used are Hermite normal forms and their unimodular multipliers. With the same line of ideas, a complete solution to the scaling symmetry reduction of a polynomial system is also presented.

This is joint work with George Labahn (University of Waterloo, Canada).

6.2.7. Scaling invariants and symmetry reduction of dynamical systems

Participant: Evelyne Hubert.

The motivation for this subject is to offer an algorithmic scheme for reducing the number of parameters in physical, chemical or biological models. This comes as a special case of a symmetry reduction scheme that can be fully realized by linear algebra over the integers. See http://hal.inria.fr/hal-00668882. We provide there the algebraic determination of the scaling symmetry of a dynamical system and an complete explicit symmetry reduction scheme with polynomial complexity.

This is joint work with George Labahn (University of Waterloo, Canada).

6.2.8. A computational approach to the discriminant of homogeneous polynomials

Participant: Laurent Busé.

In this work, the discriminant of homogeneous polynomials is studied in two particular cases: a single homogeneous polynomial and a collection of $n-1$ homogeneous polynomials in n variables. In these two cases, the discriminant is defined over a large class of coefficient rings by means of the resultant. Many formal properties and computational rules are provided and the geometric interpretation of the discriminant is investigated over a general coefficient ring, typically a domain.

This work is done in collaboration with Jean-Pierre Jouanolou (University of Strasbourg). A preprint is available at http://hal.inria.fr/hal-00747930/en/.

6.2.9. Intersection between rational curves and surfaces by means of matrix representations

Participant: Laurent Busé.

In [37], we propose a survey of matrix representations for parameterized curves and surfaces. Illustrations of the properties of these representations are given for intersection problems. In particular, we focus on the ray/surface intersection which is an important step in ray-tracing algorithms.

6.2.10. A root isolation algorithm for sparse univariate polynomials

Participant: André Galligo.

In [38], we consider a univariate polynomial f with real coefficients having a high degree N but a rather small number $d+1$ of monomials, with $d \ll N$. Such a sparse polynomial has a number of real roots smaller or equal to d. Our target is to find for each real root of f an interval isolating this root from the others. The usual subdivision methods, relying either on Sturm sequences or Moebius transform followed by Descartes's rule of signs, destruct the sparse structure. Our approach relies on the generalized Budan-Fourier theorem of Coste, Lajous, Lombardi, Roy and the techniques developed in some previous works of Galligo. To such a f is associated a set of $d+1\mathbb{F}$-derivatives. The Budan-Fourier function $V_f(x)$ counts the sign changes in the sequence of $\mathbb{F}$-derivatives of f evaluated at x. The values at which this function jumps are called the $\mathbb{F}$-virtual roots of f. These include the real roots of f. We also consider the augmented $\mathbb{F}$-virtual roots of f and introduce a genericity property which eases our study. We present a real root isolation method and an algorithm which has been implemented in Maple. We rely on an improved generalized Budan-Fourier count applied to both the input polynomial and its reciprocal, together with Newton like approximation steps.

This is a joint work with Maria Emilia Alonso (University of Madrid).

6.2.11. Deformation of roots of polynomials via fractional derivatives

Participant: André Galligo.

In [26], we first recall the main features of Fractional calculus. In the expression of fractional derivatives of a real polynomial $f(x)$, we view the order of differentiation q as a new indeterminate; then we define a new bivariate polynomial $Pf(x,q)$. For $0 \leq q \leq 1$, $Pf(x,q)$ defines a homotopy between the polynomials $f(x)$ and $xf'(x)$. Iterating this construction, we associate to $f(x)$ a plane spline curve, called the stem of f. Stems of classic random polynomials exhibits intriguing patterns; moreover in the complex plane $Pf(x,q)$ creates an unexpected correspondence between the complex roots and the critical points of $f(x)$. We propose 3 conjectures to describe and explain these phenomena. Illustrations are provided relying on the Computer Algebra System Maple.

GALLIUM Project-Team

6. New Results

6.1. Language design and type systems

6.1.1. The Mezzo programming language

Participants: Jonathan Protzenko, François Pottier.

In the past ten years, the type systems community and the separation logic community, among others, have developed highly expressive formalisms for describing ownership policies and controlling side effects in imperative programming languages. In spite of this extensive knowledge, it remains very difficult to come up with a programming language design that is simple, effective (it actually controls side effects!) and expressive (it does not force programmers to alter the design of their data structures and algorithms).

The Mezzo programming language, formerly known as HaMLet, aims to bring new answers to these questions.

We have come up with a solid design for the programming language: many features of the language have been reworked or consolidated this year, and we believe we strike a good balance between expressiveness and complexity. We wrote several flagship examples that illustrate the gains offered by Mezzo, as well as two (yet unpublished) papers discussing the design of the language. Jonathan Protzenko implemented a prototype type-checker; although it is not perfect yet, several non-trivial examples are successfully type-checked.

The current state of the Mezzo programming language is best described in [40]; a former version of this document can be found as [39].

François Pottier wrote a formal definition of (a slightly lower-level variant of) Mezzo, and proved that Mezzo is type-safe: that is, well-typed programs cannot crash (but they can stop abruptly if a run-time check fails). The proof, which is about 15,000 lines, has been machine-checked using Coq. A paper that describes this work is in preparation.

This work was facilitated by Pottier's experience with a similar previous proof. In particular, out of the above 15,000 lines, about 2,000 lines correspond to a re-usable library for working with de Bruijn indices, and about 3,000 lines correspond to a re-usable formalisation of "monotonic separation algebras", which help reason about resources (memory, time, knowledge, ...) and how they evolve over time. These libraries have not yet been fully documented and released; this might be done in the future.

6.1.2. Coercion abstraction

Participants: Julien Cretin, Didier Rémy.

Expressive type systems often allow non trivial conversions between types, which may lead to complex, challenging, and sometimes ad hoc type systems. Such examples are the extension of System F with type equalities to model GADTs and type families of Haskell, or the extension of System F with explicit contracts. A useful technique to simplify the meta-theoretical studies of such systems is to make type conversions explicit as "coercions" inside terms.

Following a general approach to coercions based on System F, we introduced a language F-iota with abstraction over coercions and where all type transformations are represented as coercions. The main difficulty is dealing with coercion abstraction, as abstract coercions whose types are uninhabited cannot be erased at run-time. We proposed a restriction, called parametric F-iota, that ensures erasability of all coercions by construction. This work was presented at the POPL conference in January [22].

We extended parametric F-iota with non-interleaved positive recursive types and with erasable isomorphisms. We generalized the presentation of the language viewing coercions as conversions between typings (pairs of a typing environment and a type) rather than between types. An extended version with full proofs will be submitted for journal publication.

We also studied a more liberal version of F-iota where coercion inhabitation is no more ensured by construction (which limits expressiveness), but instead by providing coercion witnesses in source terms. This extension requires pushing abstract coercions under redexes so that they do not block the reduction. As a consequence, coercions cannot be reified in System F, and we need a direct proof of termination of iota-reduction. We completed one such proof based on reducibility candidates.

6.1.3. Ambivalent types for principal type inference with GADTs

Participants: Jacques Garrigue [Nagoya University], Didier Rémy.

Type inference for Generalized Abstract Data Types (GADTs) is always a matter of compromise because it is inherently non monotone: assuming more specific types for GADTs may ensure more invariants, which in turn may result in more general types. Moreover, even when types of GADTs parameters are explicitly given, they introduce equalities between types, which makes them inter-convertible but with a limited scope. This may then creates an ambiguity when leaving the scope of the equation: which representative should be used for the equivalent forms? Ideally, one should use a type disjunction, but this is not allowed—for good reasons. Hence, to avoid arbitrary choices, these situations must be rejected, forcing the user to add more annotations to resolve ambiguities.

We proposed a new approach to type inference with GADTs. While some uses of equations are unavoidable and create real ambiguities, others are gratuitous and create artificial ambiguities, To distinguish between the two, we introduced *ambivalent types*: a way to trace types that have been obtained by an unavoidable use of an equation. We then redefined ambiguities so that only ambivalent types become ambiguous and should be rejected or resolved by a programmer annotation.

Interestingly, the solution is fully compatible with unification-based type inference algorithms used in ML dialects. The work was presented at the ML workshop [31] and implemented in the latest version 4.00 of OCaml.

6.1.4. GADTs and Subtyping

Participants: Gabriel Scherer, Didier Rémy.

Following the addition of GADTs to the OCaml language in version 4.00 released this year, we studied the theoretical underpinnings of variance subtyping for GADTs. The question is to decide which variances should be accepted for a GADT-style type declaration that includes type equality constraints in constructor types. This question exposes a new notion of decomposability and unexpected tensions in the design of a subtyping relation. Our formalization partially reuses earlier work by François Pottier and Vincent Simonet [54]. It was presented at the ML Workshop [33]. An extended version including full proofs is available as a technical report [38] and was submitted for presentation at a conference.

6.1.5. Singleton types for code inference

Participants: Gabriel Scherer, Didier Rémy.

Inspired by tangent aspects of the PhD work of Julien Cretin, we investigated the use of singleton types for code inference. If we can prove that a type contains, in a suitably restricted pure lambda-calculus, a unique inhabitant modulo program equivalence, the compiler can infer the code of this inhabitant. This opens the way to type-directed description of boilerplate code, through type inference of finer-grained type annotations. The preliminary results seem encouraging, both on the theoretical side (identifying general situations for type-directed programming) and the practical side (mining existing OCaml code for usage situations).

6.1.6. Programming with names and binders

Participants: Nicolas Pouillard, François Pottier.

Following Nicolas Pouillard's Ph.D. defense in January 2012 [11], Nicolas Pouillard and François Pottier produced a unified presentation of Pouillard's approach to programming with abstract syntax, in the form of a paper that was published in the Journal of Functional Programming [16].

6.1.7. A type-and-capability calculus with hidden state

Participant: François Pottier.

During the year 2010, François Pottier developed a machine-checked proof of an expressive type-and-capability system, which can be used to type-check and prove properties of imperative ML programs. The proof is carried out in Coq and takes up roughly 20,000 lines of code. In the first half of 2011, François Pottier wrote a paper that describes the system and its proof in detail. This paper was published, after a revision, in 2012 [15].

6.2. Formal verification of compilers and static analyses

6.2.1. The CompCert verified C compiler

Participants: Xavier Leroy, Sandrine Blazy [project-team Celtique], Jacques-Henri Jourdan, Valentin Robert.

In the context of our work on compiler verification (see section 3.3.1), since 2005 we have been developing and formally verifying a moderately-optimizing compiler for a large subset of the C programming language, generating assembly code for the PowerPC, ARM, and x86 architectures [5]. This compiler comprises a back-end part, translating the Cminor intermediate language to PowerPC assembly and reusable for source languages other than C [4], and a front-end translating the CompCert C subset of C to Cminor. The compiler is mostly written within the specification language of the Coq proof assistant, from which Coq's extraction facility generates executable Caml code. The compiler comes with a 50000-line, machine-checked Coq proof of semantic preservation establishing that the generated assembly code executes exactly as prescribed by the semantics of the source C program.

The two major novelties of CompCert this year are described separately: verification of floating-point arithmetic (section 6.2.2) and a posteriori validation of assembly and linking (section 6.2.3). Other improvements to CompCert include:

- The meaning of "volatile" memory accesses is now fully specified in the semantics of the CompCert C source language. Their translation to built-in function invocations, previously part of the unverified pre-front-end part of CompCert, is now proved correct.
- CompCert C now natively supports assignment between composite types (structs or unions), passing composite types by value as function parameters, and other instances of using composites as r-values, with the exception of returning composites by value from a function.
- A new pass was added to the compiler to perform inlining of functions. Its correctness proof raised interesting challenges to properly relate the (widely different) call stacks of the program before and after inlining.
- The constant propagation optimization is now able to propagate the initial values of global variables declared `const`.
- The common subexpression elimination (CSE) optimization was improved so as to eliminate more redundant memory loads.

Two versions of the CompCert development were publicly released, integrating these improvements: versions 1.10 in March and 1.11 in July. We also wrote a 50-page user's manual [37] and a technical report on the CompCert memory model [35].

In parallel, we continued our collaboration with Jean Souyris, Ricardo Bedin França and Denis Favre-Felix at Airbus. They are conducting an experimental evaluation of CompCert's usability for avionics software, and studying the regulatory issues (DO-178 certification) surrounding the potential use of CompCert in this context. Preliminary results were presented at the 2012 Embedded Real-Time Software and Systems conference (ERTS'12) [29].

6.2.2. Formalization of floating-point arithmetic in Compcert

Participants: Sylvie Boldo [project-team Toccata], Jacques-Henri Jourdan, Xavier Leroy, Guillaume Melquiond [project-team Toccata].

The aim of this research theme was to formalize the semantics and compilation of floating-point arithmetic in the CompCert compiler. Prior to this work, floating-point arithmetic was axiomatized in the Coq proof of CompCert, then mapped to OCaml's floating-point operations during extraction. This approach was prone to errors and fails to formally guarantee conformance to the IEEE-754 standard for floating-point arithmetic.

To remedy this situation, Jacques-Henri Jourdan replaced this axiomatization by a fully-formal Coq development, building on the Coq formalization of IEEE-754 arithmetic provided by the Flocq library. Sylvie Boldo and Guillaume Melquiond, authors of Flocq, adapted their library to the needs of this development. The new formalization of floating-point arithmetic is used throughout CompCert: to give semantics to FP computations in the source, intermediate and target (assembly) languages; to perform correct compile-time FP evaluations during constant propagation; to prove the correctness of code generation scheme for conversions between integers and FP numbers; and to parse FP literals with correct rounding.

A paper describing this work is accepted for presentation at the forthcoming ARITH 2013 conference [20].

6.2.3. *Validation of assembly and linking*

Participants: Valentin Robert, Xavier Leroy.

Valentin Robert designed and implemented a validation tool for the assembly and linking phases of the CompCert C compiler. These passes are not formally verified and call into off-the-shelf assemblers and linkers. The `cchecklink` tool of Valentin Robert improves the confidence that end-users can have in these passes by validating *a posteriori* their operation. The tool takes as inputs the PowerPC/ELF executable produced by the linker, as well as the abstract syntax trees for assembly files produced by the formally-verified part of CompCert. It then proceeds to establish a correspondence between the two sets of inputs, via a thorough structural analysis on the ELF executable, light disassembling of the machine code, expansion of CompCert's macro-asm instructions, and propagation of constraints over symbolic names. The tool produces detailed diagnostics if any discrepancies are found.

6.2.4. *Improving CompCert's reusability for verification tools*

Participants: Xavier Leroy, Jacques-Henri Jourdan, Andrew Appel [Princeton University], Sandrine Blazy [project-team Celtique], David Pichardie [project-team Celtique].

Several ongoing projects focus on proving the soundness of verification tools that reuse parts of the CompCert development, namely some of the intermediate languages, their formal semantics, and the CompCert passes that produce these intermediate forms. This is the case for the Verasco ANR project, which focuses on the proof of a static analyzer based on abstract interpretation, and for the Verified Software Toolchain (VST) project, led by Andrew Appel at Princeton University, which develops a concurrent separation logic embedded in Coq. However, the CompCert intermediate languages, currently designed to fit the needs of a compiler, are not perfectly suited to static analysis and deductive verification.

To improve the reusability of CompCert's Clight language in the Verasco and VST projects, Xavier Leroy is currently revising the CompCert C front-end passes so that function-local C variables whose address is never taken are pulled out of memory and replaced by nonadressable temporary variables. The resulting Clight intermediate form is much easier to analyze or prove correct, as temporary variables cannot suffer from aliasing problems.

Likewise, Sandrine Blazy, Jacques-Henri Jourdan, Xavier Leroy and David Pichardie designed a variant of CompCert's RTL intermediate language, called CFG. Like RTL, CFG represents the flow of control by a graph; unlike RTL, CFG is independent of the target processor, and supports complex expressions instead of 3-address code. These features of CFG make it a better target for static analysis, both non-relational (e.g. David Pichardie's certified interval analysis) and relational. Jacques-Henri Jourdan implemented and proved correct a compilation pass that produces CFG code from the Cminor intermediate language of CompCert.

6.2.5. *Formal verification of hardware synthesis*

Participants: Thomas Braibant, Adam Chlipala [MIT].

Verification of hardware designs has been thoroughly investigated, and yet, obtaining provably correct hardware of significant complexity is usually considered challenging and time-consuming. Hardware synthesis aims to raise the level of description of circuits, reducing the effort necessary to produce them.

This yields two opportunities for formal verification: a first option is to verify (part of) the hardware compiler; a second option is to study to what extent these higher-level design are amenable to formal proof.

During a visit at MIT, Thomas Braibant worked on the implementation and proof of correctness of a prototype hardware compiler in Coq, under Adam Chlipala's supervision. This compiler produces descriptions of circuits in RTL style from a high-level description language inspired by BlueSpec. After joining Gallium, Thomas Braibant continued working part time on this subject, finishing the proof of this compiler, and implementing a few hardware designs of mild complexity. This work was presented at the 2012 Coq Workshop [30] and will be submitted to a conference in 2013.

6.2.6. A formally-verified alias analysis

Participants: Valentin Robert, Xavier Leroy.

Valentin Robert improved the verified static analysis for pointers and non-aliasing that he initiated in 2011 during his Master's internship supervised by Xavier Leroy. This alias analysis is intraprocedural and flow-sensitive, and follows the "points-to" approach of Andersen [41]. An originality of this alias analysis is that it is conducted over the RTL intermediate language of the CompCert compiler: since RTL is essentially untyped, the traditional approaches to field sensitivity do not apply, and are replaced by a simple but effective tracking of the numerical offsets of pointers with respect to their base memory blocks. The soundness of this alias analysis is proved against the operational semantics of RTL using the Coq proof assistant and techniques inspired from abstract interpretation. A paper describing the analysis and its soundness proof was presented at the CPP 2012 conference [26].

6.3. The OCaml language and system

6.3.1. The OCaml system

Participants: Xavier Clerc [team SED], Damien Doligez, Alain Frisch [Lexifi SAS], Jacques Garrigue [University of Nagoya], Fabrice Le Fessant [Inria Saclay and OCamlPro start-up company], Jacques Le Normand [Lexifi SAS], Xavier Leroy.

This year, we released versions 4.00.0 and 4.00.1 of the OCaml system. Version 4.00.0 (released in July) is a major release that fixes about 150 reported bugs and 4 unreported bugs, and introduces 57 new features suggested by users. Version 4.00.1 (released in October) is a bug-fix release that fixes 3 major and 20 minor bugs. Damien Doligez acted as release manager for both versions.

The major innovation in OCaml 4.00 is support for generalized algebraic datatypes (GADTs). These non-uniform datatype definitions enable programmers to express some invariants over data structures, and the OCaml type-checker to enforce these invariants. They also support interesting ways of reflecting types into run-time values. GADTs are found in proof assistants such as Coq and in functional languages such as Agda and Haskell. Their integration in OCaml raised delicate issues of partial type inference and principality of inferred types, to which Jacques Garrigue and Jacques Le Normand provided original solutions [45].

Other features of this release include:

- Lightweight notations to facilitate the use of first-class modules.
- Better reporting of type errors.
- Improvements in native-code generation.
- Performance and security improvements in the hashing primitive and hash tables.
- New warnings for unused code (variables, record fields, etc.)
- A new back-end for the ARM architecture.

6.3.2. Namespaces for OCaml

Participants: Gabriel Scherer, Didier Rémy, Fabrice Le Fessant [Inria Saclay].

As part of an ongoing discussion among members of the OCaml Consortium, we investigated the formal aspects of "namespaces" and their putative status in the OCaml language. Namespaces aim at providing OCaml programmers with efficient ways to manage and structure the names of compilation units, in contrast with the flat, global space of compilation units provided today in OCaml. This formalization provides scientific support to ongoing design and engineering discussions. It was presented at the December 2011 IFIP 2.8 working group on functional programming, and at the December 2012 meeting of the OCaml Consortium.

6.4. Software specification and verification

6.4.1. Tools for TLA+

Participants: Damien Doligez, Leslie Lamport [Microsoft Research], Stephan Merz [EPI VeriDis], Tomer Libal [Microsoft Research-Inria Joint Centre], Hernán Vanzetto [Microsoft Research-Inria Joint Centre].

Damien Doligez is head of the "Tools for Proofs" team in the Microsoft-Inria Joint Centre. The aim of this team is to extend the TLA+ language with a formal language for hierarchical proofs, formalizing the ideas in [47], and to build tools for writing TLA+ specifications and mechanically checking the corresponding formal proofs.

This year, the TLA+ project released two new versions (in January and in November) of the TLA+ tools: the GUI-based TLA Toolbox and the TLA+ Proof System, an environment for writing and checking TLA+ proofs. This environment is described in a paper presented at the 2012 symposium on Formal Methods [21]. The January release (version 1.0 of TLAPS and 1.4.1 of Toolbox) added support for back-ends based on SMT provers (CVC3, Z3, Yices, VeriT), which dramatically extends the range of proof obligations that the system can discharge automatically. The November release includes many bug-fixes and performance improvements.

We have also improved the theoretical design of the proof language with respect to temporal properties. This design will be implemented in TLAPS in the near future.

Web site: http://tlaplus.net/

6.4.2. The Zenon automatic theorem prover

Participants: Damien Doligez, David Delahaye [CNAM], Mélanie Jacquel [CNAM].

Damien Doligez continued the development of Zenon, a tableau-based prover for first-order logic with equality and theory-specific extensions. Version 0.7.1 of Zenon was released in May.

David Delahaye and Mélanie Jacquel designed and implemented (with some help from Damien Doligez) an extension of Zenon called SuperZenon, based on the Superdeduction framework of Brauner, Houtmann, and Kirchner [43].

Both Zenon and SuperZenon entered the CASC theorem-proving contest, where, as expected, SuperZenon did much better than Zenon.

6.4.3. Hybrid contract checking via symbolic simplification

Participant: Na Xu.

Program errors are hard to detect or prove absent. Allowing programmers to write formal and precise specifications, especially in the form of contracts, is one popular approach to program verification and error discovery. Na Xu formalizes and implements a hybrid contract checker for a pure subset of OCaml. The key technique we use is symbolic simplification, which makes integrating static and dynamic contract checking easy and effective. This technique statically verifies that a function satisfies its contract or blames the function violating the contract. When a contract satisfaction is undecidable, it leaves residual code for dynamic contract checking.

A paper describing this result is published in the proceeding of the PEPM'2012 conference [27]. An extended version of this paper will appear in the journal Higher-Order and Symbolic Computation. Na Xu implemented this approach in a prototype based on the OCaml 3.12.1 compiler and experimented with nontrivial examples such as sorting algorithms and balancing AVL trees (see http://gallium.inria.fr/~naxu/research/hcc.html).

6.4.4. Probabilistic contracts for component-based design

Participants: Na Xu, Gregor Goessler [project-team POPART], Alain Girault [project-team POPART].

We define a framework of probabilistic contracts for constructing component-based embedded systems, based on the formalism of discrete-time Interactive Markov Chains. A contract specifies the assumptions a component makes on its context and the guarantees it provides. Probabilistic transitions represent allowed uncertainty in the component behavior, for instance, to model internal choice or reliability. Action transitions are used to model non-deterministic behavior and communication between components. An interaction model specifies how components interact with each other.

We provide the ingredients for a component-based design flow, including (1) contract satisfaction and refinement, (2) parallel composition of contracts over disjoint, interacting components, and (3) conjunction of contracts describing different requirements over the same component. Compositional design is enabled by congruence of refinement. A paper describing the details of this result is published in the journal Formal Methods in System Design [14].

GEOMETRICA Project-Team

6. New Results

6.1. Mesh Generation and Geometry Processing

6.1.1. New bounds on the size of optimal meshes

Participant: Donald Sheehy.

The theory of optimal size meshes gives a method for analyzing the output size (number of simplices) of a Delaunay refinement mesh in terms of the integral of a sizing function over the input domain. The input points define a maximal such sizing function called the feature size. This work aims to find a way to bound the feature size integral in terms of an easy to compute property of a suitable ordering of the point set. The key idea is to consider the pacing of an ordered point set, a measure of the rate of change in the feature size as points are added one at a time. In previous work, Miller et al. showed that if an ordered point set has pacing ϕ, then the number of vertices in an optimal mesh will be $O(\phi^d n)$, where d is the input dimension. We give a new analysis of this integral showing that the output size is only $\Theta(n + n\log\phi)$. The new analysis tightens bounds from several previous results and provides matching lower bounds. Moreover, it precisely characterizes inputs that yield outputs of size $O(n)$ [20].

6.1.2. State of the art in quad meshing

Participant: David Bommes.

Triangle meshes have been nearly ubiquitous in computer graphics, and a large body of data structures and geometry processing algorithms based on them has been developed in the literature. At the same time, quadrilateral meshes, especially semi-regular ones, have advantages for many applications, and significant progress was made in quadrilateral mesh generation and processing during the last several years. In this work, we discuss the advantages and problems of techniques operating on quadrilateral meshes, including surface analysis and mesh quality, simplification, adaptive refinement, alignment with features, parametrization, and remeshing [23].

6.1.3. Meshing the hyperbolic octagon

Participants: Mathieu Schmitt, Monique Teillaud.

We propose a practical method to compute a mesh of the octagon, in the Poincaré disk, that respects its symmetries. This is obtained by meshing the Schwartz triangle $T(8,3,2)$ and applying relevant hyperbolic symmetries (ie., Euclidean reflexions or inversions). The implementation is based on CGAL 2D meshes and on the ongoing implementation on CGAL hyperbolic Delaunay triangulations [44]. Further work will include solving robutsness issues and generalizing the method to any Schwartz triangle [62].

6.1.4. Index-based data structure for 3D polytopal complexes

Participant: David Bommes.

OpenVolumeMesh is a data structure which is able to represent heterogeneous 3-dimensional polytopal cell complexes and is general enough to also represent non-manifolds without incurring undue overhead [30]. Extending the idea of half-edge based data structures for two-manifold surface meshes, all faces, i.e. the two-dimensional entities of a mesh, are represented by a pair of oriented half-faces. The concept of using directed half-entities enables inducing an orientation to the meshes in an intuitive and easy to use manner. We pursue the idea of encoding connectivity by storing first-order top-down incidence relations per entity, i.e. for each entity of dimension d, a list of links to the respective incident entities is stored. For instance, each half-face as well as its orientation is uniquely determined by a tuple of links to its incident half-edges or each 3D cell by the set of incident half-faces. This representation allows for handling non-manifolds as well as mixed-dimensional mesh configurations. No entity is duplicated according to its valence, instead, it is shared by all incident entities in order to reduce memory consumption. Furthermore, an array-based storage layout is used in combination with direct index-based access. This guarantees constant access time to the entities of a mesh. Although bottom-up incidence relations are implied by the top-down incidences, our data structure provides the option to explicitly generate and cache them in a transparent manner. This allows for accelerated navigation in the local neighbor- hood of an entity. We provide an open-source and platform-independent implementation of the proposed data structure written in C++ using dynamic typing paradigms. The library is equipped with a set of STL compliant iterators, a generic property system to dynamically attach properties to all entities at run-time, and a serializer/deserializer supporting a simple file format. Due to its similarity to the OpenMesh data structure, it is easy to use, in particular for those familiar with OpenMesh. Since the presented data structure is compact, intuitive, and efficient, it is suitable for a variety of applications, such as meshing, visualization, and numerical analysis. OpenVolumeMesh is open-source software licensed under the terms of the LGPL [29].

6.1.5. Editable SQuad representation for triangle meshes

Participant: Olivier Devillers.

In collaboration with Luca Castelli Aleardi (LIX, Palaiseau) and Jarek Rossignac (Georgia Tech).

We consider the problem of designing space efficient solutions for representing the connectivity information of manifold triangle meshes. Most mesh data structures are quite redundant, storing a large amount of information in order to efficiently support mesh traversal operators. Several compact data structures have been proposed to reduce storage cost while supporting constant-time mesh traversal. Some recent solutions are based on a global re-ordering approach, which allows to implicitly encode a map between vertices and faces. Unfortunately, these compact representations do not support efficient updates, because local connectivity changes (such as edge-contractions, edge-flips or vertex insertions) require re-ordering the entire mesh. Our main contribution is to propose a new way of designing compact data structures which can be dynamically maintained. In our solution, we push further the limits of the re-ordering approaches: the main novelty is to allow to re-order vertex data (such as vertex coordinates), and to exploit this vertex permutation to easily maintain the connectivity under local changes. We describe a new class of data structures, called Editable SQuad (ESQ), offering the same navigational and storage performance as previous works, while supporting local editing in amortized constant time. As far as we know, our solution provides the most compact dynamic data structure for triangle meshes. We propose a linear-time and linear-space construction algorithm, and provide worst-case bounds for storage and time cost [25].

6.1.6. Surface reconstruction through point set structuring

Participants: Pierre Alliez, Florent Lafarge.

We present a method for reconstructing surfaces from point sets. The main novelty lies into a structure-preserving approach where the input point set is first consolidated by structuring and resampling the planar components, before reconstructing the surface from both the consolidated components and the unstructured points. The final surface is obtained through solving a graph-cut problem formulated on the 3D Delaunay triangulation of the structured point set where the tetrahedra are labeled as inside or outside cells. Structuring facilitates the surface reconstruction as the point set is substantially reduced and the points are enriched with structural meaning related to adjacency between primitives. Our approach departs from the common dichotomy between smooth/piecewise-smooth and primitive-based representations by gracefully combining canonical parts from detected primitives and free-form parts of the inferred shape. Our experiments on a variety of inputs illustrate the potential of our approach in terms of robustness, flexibility and efficiency [59].

6.1.7. Feature-preserving surface reconstruction and simplification from defect-laden point sets

Participants: Pierre Alliez, David Cohen-Steiner, Julie Digne.

In collaboration with Fernando de Goes and Mathieu Desbrun from Caltech.

We introduce a robust and feature-capturing surface reconstruction and simplification method that turns an input point set into a low triangle-count simplicial complex. Our approach starts with a (possibly non-manifold) simplicial complex filtered from a 3D Delaunay triangulation of the input points. This initial approximation is iteratively simplified based on an error metric that measures, through optimal transport, the distance between the input points and the current simplicial complex, both seen as mass distributions. Our approach is shown to exhibit both robustness to noise and outliers, as well as preservation of sharp features and boundaries (Figure 1). Our new feature-sensitive metric between point sets and triangle meshes can also be used as a post-processing tool that, from the smooth output of a reconstruction method, recovers sharp features and boundaries present in the initial point set [58].

6.1.8. Similarity based filtering of point clouds

Participant: Julie Digne.

Denoising surfaces is a crucial step in the surface processing pipeline. This is even more challenging when no underlying structure of the surface is known, that is when the surface is represented as a set of unorganized points. We introduce a denoising method based on *local similarities*. The contributions are threefold: first, we do not denoise directly the point positions but use a low/high frequency decomposition and denoise only the high frequency. Second, we introduce a local surface parameterization which is proved stable. Finally, this method works directly on point clouds, thus avoiding building a mesh of a noisy surface which is a difficult problem. Our approach is based on denoising a height vector field by comparing the neighborhood of the point with neighborhoods of other points on the surface (Figure 2). It falls into the non-local denoising framework that has been extensively used in image processing, but extends it to unorganized point clouds [26].

6.1.9. Progressive compression of manifold polygon meshes

Participant: Pierre Alliez.

In collaboration with Adrien Maglo, Clément Courbet and Céline Hudelot from Ecole Centrale Paris.

We present a new algorithm for the progressive compression of surface polygon meshes. The input surface is decimated by several traversals that generate successive levels of detail through a specific patch decimation operator which combines vertex removal and local remeshing. This operator encodes the mesh connectivity through a transformation that generates two lists of Boolean symbols during face and edge removals. The geometry is encoded with a barycentric error prediction of the removed vertex coordinates. In order to further reduce the size of the geometry and connectivity data, we propose a curvature prediction method and a connectivity prediction scheme based on the mesh geometry. We also include two methods that improve the rate-distortion performance: a wavelet formulation with a lifting scheme and an adaptive quantization technique. Experimental results demonstrate the effectiveness of our approach in terms of compression rates and rate-distortion performance. Our approach compares favorably to compression schemes specialized to triangle meshes [31].

6.2. Topological and Geometric Inference

6.2.1. Homological reconstruction and simplification in $\mathbb{R}^3$

Participants: Olivier Devillers, Marc Glisse.

In collaboration with Dominique Attali (Gipsa-lab), Ulrich Bauer (Göttingen Univ.), and André Lieutier (Dassault Systèmes).

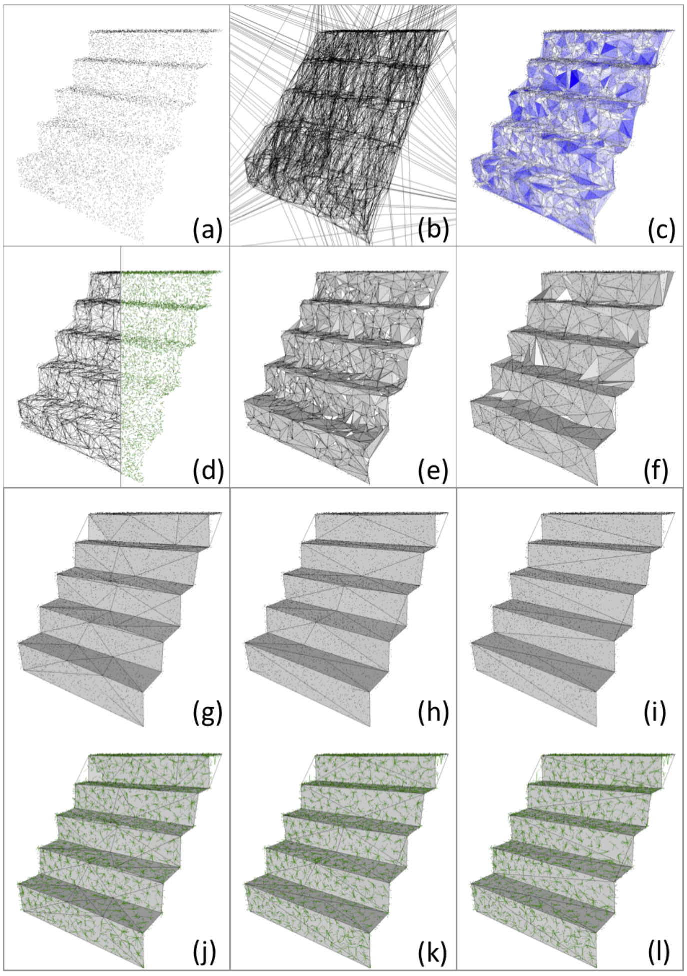


Figure 1. Steps of our algorithm: (a) Initial point set; (b) 3D Delaunay triangulation of a random subset containing 10% of the input points; (c) Initial simplicial complex constructed from facets of the 3D triangulation with non-zero measure; (d) Initial transport plan assigning point samples to bin centroids (green arrows); (e-f) Intermediary decimation steps; (g-i) Reconstruction with 100, 50, and 22 vertices, respectively; (j-l) Final transport plan with 100, 50, and 22 vertices, respectively.

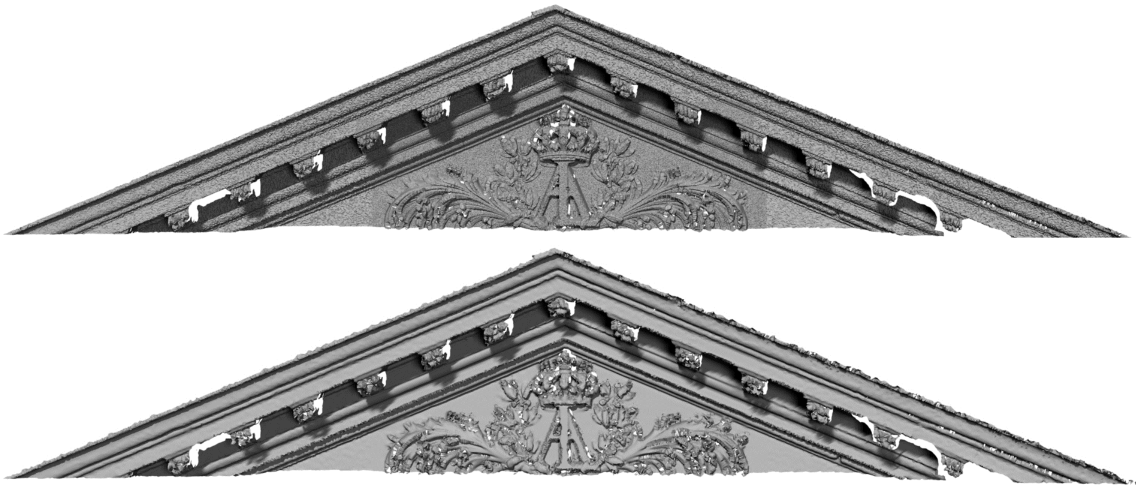


Figure 2. Similarity-based denoising. Top: input point set. Bottom: point set after denoising.

We consider the problem of deciding whether the persistent homology group of a simplicial pair (K, L) can be realized as the homology $H_*(X)$ of some space X with $L \subset X \subset K$. We show that this problem is NP-complete even if K is embedded in $\mathbb{R}^3$.

As a consequence, we show that it is NP-hard to simplify level and sublevel sets of scalar functions on $\mathbb{S}^3$ within a given tolerance constraint. This problem has relevance to the visualization of medical images by isosurfaces. We also show an implication to the theory of well groups of scalar functions: not every well group can be realized by some level set, and deciding whether a well group can be realized is NP-complete [43].

6.2.2. *The structure and stability of persistence modules*

Participants: Frédéric Chazal, Marc Glisse, Steve Oudot.

In collaboration with Vin de Silva (Pomona College)

We give a self-contained treatment of the theory of persistence modules indexed over the real line. We give new proofs of the standard results. Persistence diagrams are constructed using measure theory. Linear algebra lemmas are simplified using a new notation for calculations on quiver representations. We show that the stringent finiteness conditions required by traditional methods are not necessary to prove the existence and stability of the persistence diagram. We introduce weaker hypotheses for taming persistence modules, which are met in practice and are strong enough for the theory still to work. The constructions and proofs enabled by our framework are, we claim, cleaner and simpler [54].

6.2.3. *Persistence stability for geometric complexes*

Participants: Frédéric Chazal, Steve Oudot.

In collaboration with Vin de Silva (Pomona College)

We study the properties of the homology of different geometric filtered complexes (such as Vietoris–Rips, Čech and witness complexes) built on top of precompact spaces. Using recent developments in the theory of topological persistence [54] we provide simple and natural proofs of the stability of the persistent homology of such complexes with respect to the Gromov–Hausdorff distance. We also exhibit a few noteworthy properties of the homology of the Rips and Čech complexes built on top of compact spaces [53].

6.2.4. Zigzag zoology: rips zigzags for homology inference

Participants: Steve Oudot, Donald Sheehy.

For points sampled near a compact set X, the persistence barcode of the Rips filtration built from the sample contains information about the homology of X as long as X satisfies some geometric assumptions. The Rips filtration is prohibitively large, however zigzag persistence can be used to keep the size linear. We present several species of Rips-like zigzags and compare them with respect to the signal-to-noise ratio, a measure of how well the underlying homology is represented in the persistence barcode relative to the noise in the barcode at the relevant scales. Some of these Rips-like zigzags have been available as part of the Dionysus library for several years while others are new. Interestingly, we show that some species of Rips zigzags will exhibit less noise than the (non-zigzag) Rips filtration itself. Thus, the Rips zigzag can offer improvements in both size complexity and signal-to-noise ratio.

Along the way, we develop new techniques for manipulating and comparing persistence barcodes from zigzag modules. We give methods for reversing arrows and removing spaces from a zigzag. We also discuss factoring zigzags and a kind of interleaving of two zigzags that allows their barcodes to be compared. These techniques were developed to provide our theoretical analysis of the signal-to-noise ratio of Rips-like zigzags, but they are of independent interest as they apply to zigzag modules generally [60].

6.2.5. A space and time efficient implementation for computing persistent homology

Participants: Jean-Daniel Boissonnat, Clément Maria.

In collaboration with Tamal Dey (Ohio State University)

The persistent homology with Z_2-coefficients coincides with the same for cohomology because of duality. Recently, it has been observed that the cohomology based algorithms perform much better in practice than the originally proposed homology based persistence algorithm. We have implemented a cohomology based algorithm that attaches binary labels called annotations with the simplices. This algorithm fits very naturally with our recently developed data structure called simplex tree to represent simplicial complexes [49], [22]. By taking advantages of several practical tricks such as representing annotations compactly with memory words, using a union-find structure that eliminates duplicate annotation vectors, and a lazy evaluation, we save both space and time cost for computations. The complexity of the procedure, in practice, depends almost linearly on the size of the simplicial complex and on the variables related to the maximal dimension of the local homology groups we maintain during the computation, which remain small in practice. We provide a theoretical analysis as well as a detailed experimental study of our implementation. Experimental results show that our implementation performs several times better than the existing state-of-the-art software for computing persistent homology in terms of both time and memory requirements and can handle very large (several hundred million simplices in high-dimension) complexes efficiently [45].

6.2.6. Minimax rates for homology inference

Participant: Donald Sheehy.

In collaboration with Sivaraman Balakrishnan and Alessandro Rinaldo and Aarti Singh and Larry A. Wasserman (Carnegie Mellon University)

Often, high dimensional data lie close to a low-dimensional submanifold and it is of interest to understand the geometry of these submanifolds. The homology groups of a manifold are important topological invariants that provide an algebraic summary of the manifold. These groups contain rich topological information, for instance, about the connected components, holes, tunnels and sometimes the dimension of the manifold. We consider the statistical problem of estimating the homology of a manifold from noisy samples under several different noise models. We derive upper and lower bounds on the minimax risk for this problem. Our upper bounds are based on estimators which are constructed from a union of balls of appropriate radius around carefully selected points. In each case, we establish complementary lower bounds using Le Cam's lemma [15].

6.2.7. Linear-size approximations to the Vietoris-Rips filtration

Participant: Donald Sheehy.

The Vietoris-Rips filtration is a versatile tool in topological data analysis. Unfortunately, it is often too large to construct in full. We show how to construct an $O(n)$-size filtered simplicial complex on an n-point metric space such that the persistence diagram is a good approximation to that of the Vietoris-Rips filtration. The filtration can be constructed in $O(n \log n)$ time. The constants depend only on the doubling dimension of the metric space and the desired tightness of the approximation. For the first time, this makes it computationally tractable to approximate the persistence diagram of the Vietoris-Rips filtration across all scales for large data sets. Our approach uses a hierarchical net-tree to sparsify the filtration. We can either sparsify the data by throwing out points at larger scales to give a zigzag filtration, or sparsify the underlying graph by throwing out edges at larger scales to give a standard filtration. Both methods yield the same guarantees [34].

6.2.8. *A multicover nerve for geometric inference*

Participant: Donald Sheehy.

We show that filtering the barycentric decomposition of a Čech complex by the cardinality of the vertices captures precisely the topology of k-covered regions among a collection of balls for all values of k. Moreover, we relate this result to the Vietoris-Rips complex to get an approximation in terms of the persistent homology [33].

6.2.9. *Computing well diagrams for vector fields on* $\mathbb{R}^n$

Participant: Frédéric Chazal.

In collaboration with Primoz Skraba (Lubiana Univ.), Amit Patel (Rutgers Univ.)

Using topological degree theory, we present and prove correctness of a fast algorithm for computing the well diagram, a quantitative property, of a vector field on Euclidean space [17].

6.3. Data Structures and Robust Geometric Computation

6.3.1. *Straight-line graph drawing on the torus*

Participant: Olivier Devillers.

In collaboration with Luca Castelli Aleardi and Éric Fusy (LIX, Palaiseau).

We extend the notion of canonical orderings to cylindric triangulations. This allows us to extend the incremental straight-line drawing algorithm of de Fraysseix et al. to this setting. Our algorithm yields in linear time a crossing-free straight-line drawing of a cylindric triangulation T with n vertices on a regular grid $\mathbb{Z}/w\mathbb{Z} \times [0, h]$, with $w \leq 2n$ and $h \leq n(2d+1)$, where d is the (graph-) distance between the two boundaries. As a by-product, we can also obtain in linear time a crossing-free straight-line drawing of a toroidal triangulation with n vertices on a periodic regular grid $\mathbb{Z}/w\mathbb{Z} \times \mathbb{Z}/h\mathbb{Z}$, with $w \leq 2n$ and $h \leq 1 + n(2c+1)$, where c is the length of a shortest non-contractible cycle. Since $c \leq \sqrt{2n}$, the grid area is $O(n^{5/2})$ [24].

6.3.2. *Qualitative symbolic perturbation*

Participants: Olivier Devillers, Monique Teillaud.

In collaboration with Menelaos Karavelas (University of Crete).

In the literature, the generic way to address degeneracies in computational geometry is the *Symbolic Perturbation* paradigm: the input is made dependent of some parameter ε so that for ε positive and close to zero, the input is close to the original input, while at the same time, in non-degenerate position. A geometric predicate can usually be seen as the sign of some function of the input. In the symbolic perturbation paradigm, if the function evaluates to zero, the input is perturbed by a small positive ε, and the sign of the function evaluated at the perturbed input is used instead.

The usual way of using this approach is what we will call *Algebraic Symbolic Perturbation* framework. When the function to be evaluated is a polynomial of the input, its perturbed version is seen as a polynomial in ε, whose coefficients are polynomials in the input. These coefficients are evaluated by increasing degree in ε until a non-vanishing coefficient is found. The number of these coefficients can be quite large and expressing them in an easily and efficiently computable manner (e.g., factorized) may require quite some work.

We propose to address the handling of geometric degeneracies in a different way, namely by means of what we call the *Qualitative Symbolic Perturbation* framework. We no longer use a single perturbation that must remove all degeneracies, but rather a sequence of perturbations, such that the next perturbation is being used only if the previous ones have not removed the degeneracies. The new perturbation is considered as *symbolically smaller* than the previous ones. This approach allows us to use simple elementary perturbations whose effect can be analyzed and evaluated: (1) by geometric reasoning instead of algebraic development of the predicate polynomial in ε, and (2) independently of a specific algebraic formulation of the predicate.

We apply our framework to predicates used in the computation of Apollonius diagrams in 2D and 3D, as well as the computation of trapezoidal maps of circular arcs [57].

6.3.3. Covering spaces and Delaunay triangulations of the 2D flat torus

Participants: Mikhail Bogdanov, Monique Teillaud.

In collaboration with Gert Vegter (Johan Bernoulli Institute, Groningen University)

A previous algorithm was computing the Delaunay triangulation of the flat torus, by using a 9-sheeted covering space [64]. We propose a modification of the algorithm using only a 8-sheeted covering space, which allows to work with 8 periodic copies of the input points instead of 9. The main interest of our contribution is not only this result, but most of all the method itself: this new construction of covering spaces generalizes to Delaunay triangulations of surfaces of higher genus.

6.3.4. Hyperbolic Delaunay complexes and Voronoi diagrams made practical

Participants: Mikhail Bogdanov, Olivier Devillers, Monique Teillaud.

We study Delaunay complexes and Voronoi diagrams in the Poincaré ball, a confomal model of the hyperbolic space, in any dimension. We elaborate on our earlier work on the space of spheres [65], giving a detailed description of algorithms, and presenting a static and a dynamic variants. All proofs are based on geometric reasoning, they do not resort to any use of the analytic formula of the hyperbolic distance. We also study algebraic and arithmetic issues, observing that only rational computations are needed. This allows for an exact and efficient implementation in 2D. All degenerate cases are handled. The implementation will be submitted to the CGAL editorial board for future integration into the CGAL library [44].

6.3.5. The stability of Delaunay triangulations

Participants: Jean-Daniel Boissonnat, Ramsay Dyer.

In collaboration with Arijit Ghosh (Indian Statistical Institute, Kolkata, India)

We introduce a parametrized notion of genericity for Delaunay triangulations which, in particular, implies that the Delaunay simplices of δ-generic point sets are thick. Equipped with this notion, we study the stability of Delaunay triangulations under perturbations of the metric and of the vertex positions. We quantify the magnitude of the perturbations under which the Delaunay triangulation remains unchanged. We also present an algorithm that takes as input a discrete point set in $\mathbb{R}^m$, and performs a small perturbation that guarantees that the Delaunay triangulation of the resulting perturbed point set has quantifiable stability with respect to the metric and the point positions. There is also a guarantee on the quality of the simplices: they cannot be too flat. The algorithm provides an alternative tool to the weighting or refinement methods to remove poorly shaped simplices in Delaunay triangulations of arbitrary dimension, but in addition it provides a guarantee of stability for the resulting triangulation [21], [47].

6.3.6. Constructing intrinsic Delaunay triangulations of submanifolds

Participants: Jean-Daniel Boissonnat, Ramsay Dyer.

In collaboration with Arijit Ghosh (Indian Statistical Institute, Kolkata, India)

This work is the algorithmic counterpart of our previous paper [21]. We describe an algorithm to construct an intrinsic Delaunay triangulation of a smooth closed submanifold of Euclidean space. We also provide a counterexample to the results announced by Leibon and Letscher on Delaunay triangulations on Riemannian manifolds. In general the nerve of the intrinsic Voronoi diagram is not homeomorphic to the manifold. The density of the sample points alone cannot guarantee the existence of a Delaunay triangulation. To circumvent this issue, we use results established in our companion paper on the stability of Delaunay triangulations on δ-generic point sets. We establish sampling criteria which ensure that the intrinsic Delaunay complex coincides with the restricted Delaunay complex and also with the recently introduced tangential Delaunay complex. The algorithm generates a point set that meets the required criteria while the tangential complex is being constructed. In this way the computation of geodesic distances is avoided, the runtime is only linearly dependent on the ambient dimension, and the Delaunay complexes are guaranteed to be triangulations of the manifold [46].

6.3.7. Equating the witness and restricted Delaunay complexes

Participants: Jean-Daniel Boissonnat, Ramsay Dyer, Steve Oudot.

In collaboration with Arijit Ghosh (Indian Statistical Institute, Kolkata, India)

It is a well-known fact that the restricted Delaunay and witness complexes may differ when the landmark and witness sets are located on submanifolds of Rd of dimension 3 or more. Currently, the only known way of overcoming this issue consists of building some crude superset of the witness complex, and applying a greedy sliver exudation technique on this superset. Unfortunately, the construction time of the superset depends exponentially on the ambient dimension, which makes the witness complex based approach to manifold reconstruction impractical. This work provides an analysis of the reasons why the restricted Delaunay and witness complexes fail to include each other. From this, a new set of conditions naturally arises under which the two complexes are equal [37].

6.3.8. Simpler complexity analysis of random geometric structures

Participants: Olivier Devillers, Marc Glisse.

In collaboration with Xavier Goaoc (EPI VEGAS*).*

Average-case analysis of data-structures or algorithms is commonly used in computational geometry when the, more classical, worst-case analysis is deemed overly pessimistic. Since these analyses are often intricate, the models of random geometric data that can be handled are often simplistic and far from “realistic inputs”. We present a new simple scheme for the analysis of geometric structures. While this scheme only produces results up to a polylog factor, it is much simpler to apply than the classical techniques and therefore succeeds in analyzing new input distributions related to smoothed complexity analysis.

We illustrate our method on two classical structures: convex hulls and Delaunay triangulations. Specifically, we give short and elementary proofs of the classical results that n points uniformly distributed in a ball in $\mathbb{R}^d$ have a convex hull and a Delaunay triangulation of respective expected complexities $\widetilde{\Theta}(n^{\frac{d-1}{d+1}})$ and $\widetilde{\Theta}(n)$. We then prove that if we start with n points well-spread on a sphere, e.g. an (ϵ, κ)-sample of that sphere, and perturb that sample by moving each point randomly and uniformly within distance at most δ of its initial position, then the expected complexity of the convex hull of the resulting point set is $\widetilde{\Theta}\left((\sqrt{n})^{1-\frac{1}{d}} \left(\frac{1}{\sqrt[4]{\delta}}\right)^{d-\frac{1}{d}} \right)$ [55].

6.3.9. Analysis of cone vertex walk in Poisson Delaunay triangulation

Participants: Olivier Devillers, Ross Hemsley.

In collaboration with Nicolas Broutin (EPI RAP*).*

Walking strategies are a standard tool for point location in a triangulation of size n. Although often claimed to be $\Theta(\sqrt{n}\,)$ under random distribution hypotheses, this conjecture has only been formally proved by Devroye, Lemaire, and Moreau [*Comp Geom–Theor Appl*, vol. 29, 2004], in the case of the so called *straight walk* which has the very specific property that deciding whether a given (Delaunay) triangle belongs to the walk may be determined without looking at the other sites. We analyze a different walking strategy that follows vertex neighbour relations to move towards the query. We call this walk *cone vertex walk*. We prove that cone vertex walk visits $\Theta(\sqrt{n}\,)$ vertices and can be constructed in $\Theta(\sqrt{n}\,)$ time. We provide explicit bounds on the hidden constants [50].

6.3.10. The monotonicity of f-vectors of random polytopes

Participants: Olivier Devillers, Marc Glisse.

In collaboration with Xavier Goaoc and Guillaume Moroz (EPI VEGAS*) and Matthias Reitzner (Universität Osnabrück, Germany).*

Let K be a compact convex body in $\mathbb{R}^d$, let K_n be the convex hull of n points chosen uniformly and independently in K, and let $f_i(K_n)$ denote the number of i-dimensional faces of K_n.

We show that for planar convex sets, $E[f_0(K_n)]$ is increasing in n. In dimension $d \geq 3$, we prove that if $\lim_{n\to\infty} \frac{E[f_{d-1}(K_n)]}{An^c} = 1$ for some constants A and $c > 0$ then the function $n \mapsto E[f_{d-1}(K_n)]$ is increasing for n large enough. In particular, the number of facets of the convex hull of n random points distributed uniformly and independently in a smooth compact convex body is asymptotically increasing. Our proof relies on a *random sampling* argument [57].

6.3.11. Efficient Monte Carlo sampler for detecting parametric objects in large scenes

Participants: Florent Lafarge, Yannick Verdie.

Point processes have demonstrated efficiency and competitiveness when addressing object recognition problems in vision. However, simulating these mathematical models is a difficult task, especially on large scenes. Existing samplers suffer from average performances in terms of computation time and stability. We propose a new sampling procedure based on a Monte Carlo formalism. Our algorithm exploits Markovian properties of point processes to perform the sampling in parallel. This procedure is embedded into a data-driven mechanism such that the points are non-uniformly distributed in the scene. The performances of the sampler are analyzed through a set of experiments on various object recognition problems from large scenes, and through comparisons to the existing algorithms [35], [63].

6.4. Applications

6.4.1. Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation

Participant: Florent Lafarge.

We present a novel and robust method for modeling cities from 3D-point data. Our algorithm provides a more complete description than existing approaches by reconstructing simultaneously buildings, trees and topologically complex grounds. A major contribution of our work is the original way of modeling buildings which guarantees a high generalization level while having semantized and compact representations. Geometric 3D-primitives such as planes, cylinders, spheres or cones describe regular roof sections, and are combined with mesh-patches that represent irregular roof components. The various urban components interact through a non-convex energy minimization problem in which they are propagated under arrangement constraints over a planimetric map. Our approach is experimentally validated on complex buildings and large urban scenes of millions of points, and is compared to state-of-the-art methods [19].

6.4.2. The sticky geometry of the cosmic web

Participant: Monique Teillaud.

In collaboration with Johan Hidding, Rien van de Weygaert, Bernard J.T. Jones (Kapteyn Institute, Groningen University) and Gert Vegter (Johan Bernoulli Institute, Groningen University)

We highlight the application of Computational Geometry to our understanding of the formation and dynamics of the Cosmic Web. The emergence of this intricate and pervasive weblike structure of the Universe on Megaparsec scales can be approximated by a well-known equation from fluid mechanics, the Burgers' equation. The solution to this equation can be obtained from a geometrical formalism. We have extended and improved this method by invoking weighted Delaunay and Voronoi tessellations. The duality between these tessellations finds a remarkable and profound reflection in the description of physical systems in Eulerian and Lagrangian terms [28].

GRACE Team

5. New Results

5.1. Modular curves

F. Morain has been studying the theory and practice of modular curves associated with Weber's invariants. His paper ... is accepted for publication in *Acta Arithmetica*.

5.2. Computing discrete logarithms using codes

D. Augot and F. Morain have been working on the practical application of Reed–Solomon decoding to speed up discrete logarithm computations, following the work of Cheng and Wan. This work is available as a preprint [22], and a Magma implementation was written in support of the many experiments needed.

5.3. Interleaved codes and codes over rings

G. Quintin designed a decoding algorithm based on a lifting decoding scheme. He obtained a unique decoding algorithm with quasi-linear complexity in all parameters for Reed–Solomon codes over Galois rings. Using erasures, he improved the decoding radius with the same complexity. He then applied these techniques to interleaved linear codes over a finite field, and obtained a decoding algorithm that can recover more errors than half the minimum distance. This work has been presented at IEEE ISIT 2012 (Boston, USA).

5.4. Number fields codes

J.-F. Biasse and G. Quintin described an algorithm for list decoding algebraic number field codes in polynomial time in [24]. This is the first explicit procedure for decoding number field codes, whose construction were previously described by Lenstra [33] and Guruswami [32]. They rely on a new algorithm for computing the Hermite normal form of the basis of an $\mathcal{O}_K$-module due to Biasse and Fieker [31], where $\mathcal{O}_K$ is the ring of integers of a number field K. This work has been presented at IEEE ISIT 2012 (Boston, USA).

5.5. Point counting using p-adic methods

C. Gonçalvès designed a new algorithm to compute Zeta functions of cyclic covers of the projective line. This algorithm is a generalisation of the one for superelliptic curves provided by P. Gaudry and N. Gürel and has the same complexity. Moreover, optimal bounds for the precision have been proved. An alternative basis for computations has been studied and the resulting algorithm is faster, even if the asymptotic complexity is the same.

5.6. Codes and Cartier Operator

A. Couvreur proposed a new construction of codes from algebraic curves over a finite field in [25]. This class of codes is a natural geometric generalisation of classical Goppa codes. In particular, the nice equalities "$\Gamma(L, g^{q-1}) = \Gamma(L, g^q)$" satisfied by classical Goppa codes (for instance, see [30]) extend naturally to this larger class of codes. This article is to appear in *Proceeding of the American Mathematical Society*.

5.7. Quantum Codes

A. Couvreur, N. Delfosse and G. Zémor studied a construction of quantum LDPC codes proposed by McKay, Mitchison and Shokrollahi in a draft. This construction involves Cayley graphs of $GF(2)^n$. A general lower bound for the minimum distance of such codes has been found. In addition, a family of such codes whose parameters are proved to be $[[n, O(\sqrt{n}, O(\sqrt{N}))]]$ is exhibited. Notice that up to now, no construction of quantum LDPC codes is known to have a minimum distance better that $O(\sqrt{n})$. The obtained parameters beat many well–known constructions. This work has been presented at IEEE ISIT 2011 (St Petersburg, Russia), and a long version paper [26] has been submitted to an international journal.

5.8. Code-based McEliece like cryptology

A. Couvreur is working with P. Gaborit, V. Gauthier, A. Otmani, and J.-P. Tillich on distinguisher-based attacks on cryptosystems based on Generalised Reed–Solomon codes. Using the particular structure of the square of an evaluation code, they have been able to break some variants of McEliece's cryptosystem using Generalised Reed–Solomon codes, such as Wieschebrink's variant [34]. An article is in preparation.

5.9. Cyclic Codes

A. Zeh is working with A. Wachter-Zeh (University of Ulm and Institut de Recherche de Mathématique de Rennes) and Sergey Bezzateev (St. Petersburg State University of Aerospace Instrumentation) on a new bound for the minimum distance of q-ary cyclic codes [19], [18]. The connection to the BCH bound and the Hartmann–Tzeng (HT) bound was formulated explicitly. Furthermore, the bound was refined for several families of cyclic codes. We defined syndromes and formulated a Key Equation that allows an efficient decoding up to our bound with the Extended Euclidean Algorithm. It turned out that low-rate cyclic codes with small minimum distances are useful for our approach.

5.10. Iterative List Decoding

A. Zeh is working with J. S. R. Nielsen (Department of Mathematics, DTU) on an iterative list decoding algorithm for generalized Reed–Solomon codes. The method is parametrizable and allows variants of the usual list decoding approach. An article is in preparation.

LFANT Project-Team

6. New Results

6.1. Class groups and other invariants of number fields

Participants: Karim Belabas, Jean-François Biasse, Jean-Paul Cerri, Pierre Lezowski.

P. Lezowski extended J.-P. Cerri's algorithm, which was restricted to totally real number fields, to decide whether a generic number field is norm-Euclidean. His procedure allowed to find principal and non norm-Euclidean number fields of various signatures and degrees up to 8, but also to give further insight about the norm-Euclideanity of some cyclotomic fields. Besides, many new examples of generalised Euclidean and 2-stage Euclidean number fields were obtained. The article [31] will appear in *Mathematics of Computation*.

In another direction, norm-Euclidean ideal classes have been studied. They generalise the notion of norm-Euclideanity to non principal number fields. Very few such number fields were known before. A modification of the algorithm provided many new examples and allowed to complete the study of pure cubic fields equipped with a norm-Euclidean ideal class [15].

J.-F.Biasse has determined a class of number fields for which the ideal class group, the regulator, and a system of fundamental units of the maximal order can be computed in subexponential time $L(1/3, O(1))$ (whereas the best previously known algorithms have complexity $L(1/2, O(1))$). This class of number fields is analogous to the class of curves described in [10]. The article [22] has been submitted to *Mathematics of Computation*.

Assuming the GRH, Bach proved that one can calculate the residue of the Dedekind zeta function of a number field K from the knowledge of the splitting of primes $p < X$, with an error bounded explicitly in terms of X and the field discriminant. This is a crucial ingredient in all algorithms used to compute class groups and unit groups in subexponential time (under GRH). Using Weil's explicit formula, K. Belabas improved on Bach's bound, speeding up by a sizable constant factor this part of the class group algorithm. The article has been submitted to *Mathematics of Computation*.

6.2. Number and function fields

Participants: Athanasios Angelakis, Karim Belabas, Pieter Rozenhart.

In joint work with R. Scheidler and M. Jacobson, P. Rozenhart has generalized Belabas's algorithm for tabulating cubic number fields to cubic function fields [17]. This generalization required function field analogues of the Davenport-Heilbronn Theorem and of the reduction theory of binary cubic and quadratic forms. As an additional application, they have modified the tabulation algorithm to compute 3-ranks of quadratic function fields by way of a generalisation of a theorem due to Hasse. The algorithm, whose complexity is quasi-linear in the number of reduced binary cubic forms up to some upper bound X, works very well in practice. A follow-up article [35] describes how to use these results to compute 3-ranks of quadratic function fields, in particular yielding examples of unusually high 3-rank.

In 1976, Onabe discovered that, in contrast to the Neukirch–Uchida results that were proved around the same time, a number field K is not completely characterised by its absolute abelian Galois group A_K. The first examples of non-isomorphic K having isomorphic A_K were obtained on the basis of a classification by Kubota of idele class character groups in terms of their infinite families of Ulm invariants, and did not yield a description of A_K. In [21], A. Angelakis and P. Stevenhagen provide a direct "computation" of the profinite group A_K for imaginary-quadratic K, and use it to obtain many different K that all have the same minimal absolute abelian Galois group.

On March 29–April 2, 2010, a meeting was organized by J.-M. Couveignes, D. Bertrand, Ph. Boalch and P. Debes, at the Luminy CIRM (France) on geometric and differential Galois theories, witnessing the close ties these theories have woven in recent years. The volume [18] collects the proceedings of this meeting. The articles gathered in this volume cover the following topics: moduli spaces of connections, differential equations and coverings in finite characteristic, liftings, monodromy groups in their various guises (tempered fundamental group, motivic groups, generalised difference Galois groups), and arithmetic applications.

Using Galois theory of extension rings, J.-M. Couveignes, R. Lercier and T. Ezome have proposed a new pseudo-primality test in [13]. For every positive integer $k \leq logn$, this test achieves the security of k Miller-Rabin tests at the cost of $k^{1/2} + o(1)$ Miller-Rabin tests. The implementation in Magma shows that this test is competitive for primes with a few thousands digits.

6.3. Quaternion algebras

Participants: Jean-Paul Cerri, Pierre Lezowski, Aurel Page.

With J. Chaubert, J.-P. Cerri and P. Lezowski have studied whether some quaternion fields over number fields are Euclidean, that is to say whether they admit a left or right Euclidean order. In particular, they have established the complete list of totally definite and Euclidean quaternion fields over the rationals or a quadratic number field. Moreover, they have proved that every field in this list is in fact norm Euclidean. The proofs are both theoretical and algorithmic. The article [23] will appear in *International Journal of Number Theory*.

Starting with an order in a suitable quaternion algebra over a number field F with exactly one complex place, one can construct discrete subgroups of $\mathrm{PSL}_2(\mathbb{C})$. These groups, called arithmetic Kleinian groups, act properly discontinuously with finite covolume on the hyperbolic 3-space. In [34], A. Page designs an efficient algorithm which computes a fundamental domain and a presentation for such a group. It is a generalization to the dimension 3 of an algorithm of J. Voight's [44] together with a new, nondeterministic, but faster enumeration procedure. A public implementation is available in KLEINIANGROUPS (see 5.8).

6.4. Complex multiplication and modularity

Participants: Jean-Marc Couveignes, Andreas Enge, Nicolas Mascot, Aurel Page, Damien Robert.

The article by D. Lubicz and D. Robert which explains how to compute an isogeny between two abelian varieties given the kernel (but with different levels of theta structures) has been published [16]. The preprint [25] with R. Cosset and D. Robert extends these method to provide an algorithm constructing the corresponding isogeny without changing the level. This give the first algorithm allowing to compute in polynomial time an isogeny between abelian varieties, and a public implementation is available in AVISOGENIES. The drawback of this algorithm is that it needs the geometric points of the kernel. To compute an isogeny of degree ℓ^g over a finite field, working with geometric points requires to take an extension of degree up to $\ell^g - 1$, and the situation is much worse over a number field. Recently, D. Lubicz and D. Robert have explained how to compute the corresponding isogeny given only the equations of the kernel. This gives a quasi-linear algorithm (in the degree ℓ^g of the isogeny) when ℓ is congruent to 1 modulo 4.

With K. Lauter, D. Robert has worked on improving the computation of class polynomials in genus 2 by the CRT method. The main improvements come from using the above isogeny computation, both to find a maximal curve from a curve in the correct isogeny class, and to find all other maximal curves from one. Further improvements are in the endomorphism ring computation to detect if the curve is maximal, a better sieving of the primes used (and a dynamic selection of them), and the use of the CRT over the real quadratic field rather than over $\mathbb{Q}$ for the case of dihedral CM fields to find factors of the class polynomials. These results have been published at the ANTS conference [30].

With C. Ritzenthaler, Damien Robert has shown how to compute explicitly the Serre obstruction for abelian varieties isogenous to a product of three elliptic curves. This allows to find genus 3 curves with many points over a finite field. The corresponding code has been implemented in an (experimental) version of AVISOGENIES.

In [24], H. Cohen studies several methods for the numerical computation of Petersson scalar products. In particular he proves a generalisation of Haberland's formula to any subgroup of finite index G of $\Gamma = \mathrm{PSl}_2(Z)$, which gives a fast method to compute these scalar products when a Hecke eigenbasis is not necessarily available.

J.-M. Couveignes and B. Edixhoven explore in [19] the relevance of numerical methods in dealing with higher genus curves and their Jacobians. Fast exponentiation is crucial in this context as a stable substitute to Newton's method and analytic continuation. Arakelov theory provides the necessary complexity estimates.

With Reynald Lercier, J.-M. Couveignes has given in [26] a quasi-linear time randomised algorithm that on input a finite field $\mathbb{F}_q$ with q elements and a positive integer d outputs a degree d irreducible polynomial in $\mathbb{F}_q[x]$. The running time is $d^{1+o(1)} \times (\log q)^{5+o(1)}$ elementary operations. The $o(1)$ in $d^{1+o(1)}$ is a function of d that tends to zero when d tends to infinity. And the $o(1)$ in $(\log q)^{5+o(1)}$ is a function of q that tends to zero when q tends to infinity. The fastest previously known algorithm for this purpose was quadratic in the degree. The algorithm relies on the geometry of elliptic curves over finite fields (complex multiplication) and on a recent algorithm by Kedlaya and Umans for fast composition of polynomials.

In [32], N. Mascot shows how to compute modular Galois representations associated with a newform f and the coefficients of f modulo a small prime ℓ. To this end, he designs a practical variant of the complex approximation method presented in the book edited by B. Edixhoven and J.-M. Couveignes [8]. Its efficiency stems from several new ingredients. For instance, he uses fast exponentiation in the modular Jacobian instead of analytic continuations, which greatly reduces the need to compute abelian integrals, since most of the computation handles divisors. Also, he introduces an efficient way to compute arithmetically well-behaved functions on Jacobians. He illustrates the method on the newform Δ, and manages to compute for the first time the associated faithful representation modulo ℓ and the values modulo ℓ of Ramanujan's τ function at huge primes for $\ell \in \{11, 13, 17, 19\}$. In particular, he gets rid of the sign ambiguity stemming from the use of a non-faithful representation as in J. Bosman's work.

A. Enge and R. Schertz determine in [29] under which conditions singular values of multiple η-quotients of square-free level, not necessarily prime to 6, yield class invariants, that is, algebraic numbers in ring class fields of imaginary-quadratic number fields. It turns out that the singular values lie in subfields of the ring class fields of index $2^{k'-1}$ when $k' \geq 2$ primes dividing the level are ramified in the imaginary-quadratic field, which leads to faster computations of elliptic curves with prescribed complex multiplication. The result is generalised to singular values of modular functions on $X_0^+(p)$ for p prime and ramified.

With F. Morain, A. Enge has determined exhaustively under which conditions "generalised Weber functions", that is, simple quotients of η functions of not necessarily prime transformation level and not necessarily of genus 1, yield class invariants [28]. The result is a new infinite family of generators for ring class fields, usable to determine complex multiplication curves. We examine in detail which lower powers of the functions are applicable, thus saving a factor of up to 12 in the size of the class polynomials, and describe the cases in which the polynomials have integral rational instead of integral quadratic coefficients.

6.5. Elliptic curve cryptology

Participants: Jean-Marc Couveignes, Andreas Enge, Damien Robert.

With J.-G. Kammerer, J.-M. Couveignes has given in [14] an appropriate geometric method for studying and classifying encodings into elliptic curves in a cryptographic context. Such encodings were first proposed by Icart in 2009, and later on by Farashahi, Kammerer, Lercier, and Renault. But it was a little bit disappointing to see that it was no more than an application of Tartaglia's result without any geometrical explanations for the existence of such "parameterisations" of elliptic curves. Couveignes and Kammerer have filled this gap by giving exactly what can be expected from geometry: a clear explanation. Moreover, they unify all the recent "parameterisations" of elliptic curves under the same geometric point of view. The approach described in this article uses dual curves with some results coming from intersection theory. The main originality of this work is that these geometrical tools are employed to explain symbolic computations used in cryptography, that is, encoding on elliptic curves.

The survey [20], to be published in the *Handbook of Finite Fields*, presents the state of the art of the use of elliptic curves in cryptography.

6.6. Pairings

Participants: Andreas Enge, Damien Robert, Jérôme Milan.

In [27], A. Enge gives an elementary and self-contained introduction to pairings on elliptic curves over finite fields. For the first time in the literature, the three different definitions of the Weil pairing are stated correctly and proved to be equivalent using Weil reciprocity. Pairings with shorter loops, such as the ate, ate$_i$, R-ate and optimal pairings, together with their twisted variants, are presented with proofs of their bilinearity and non-degeneracy. Finally, different types of pairings are reviewed in a cryptographic context. The article can be seen as an update chapter to [42].

MARELLE Project-Team

5. New Results

5.1. Coq and SMT provers

Participants: Michaël Armand, Benjamin Grégoire, Laurent Théry.

Continuing the work of previous years, we added an extra theory to the interface between Coq and Satisfiability Modulo Theory (SMT) provers: instantiation. It is the last really needed piece to make our tactic based on SMT provers really useful to Coq users. Part of the work was to make the proof work on statements existing in the Propositional type instead of the boolean type. This requires a change in the correctness proof.

5.2. Formal proofs on Pi

Participant: Yves Bertot.

We studied the chain of definitions and proofs necessary to show that

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} \cdots$$

and removed the axiom that was left on this topic in Coq's standard library. This part re-used a past contribution of Guillaume Allais during an internship from Ecole Normale Supérieure de Lyon. We then added a study of Machin's formula to compute decimals of π.

5.3. Formal proofs on linear algebra

Participants: Guillaume Cano, Maxime Dénès, Anders Mörtberg [University of Chalmers, Sweden], Vincent Siles [University of Chalmers, Sweden], Yves Bertot.

This year we completed a work on matrix canonical forms, providing formal proofs for the following results:

- Smith normal forms of matrices on principal ideal domains are unique,
- Every matrix on a field is similar to its Frobenius normal form
- Every matrix on an algebraically closed field is similar to its Jordan normal form

We also studied techniques to combine high-level mathematical descriptions and proofs of algorithms with executable implementations. This work led to a publication at ITP'12 [10]. We are still working on extending this work to rational numbers and real algebraic numbers.

We then worked on tools to automate proofs. In the ring tactic, all elements considered must belong to the same type. We worked on extending this tactic to dependent families of types, like the type of matrices where each dimension gives rise to a different type in the family and multiplications typically concern matrices of different types, while remaining associative.

5.4. Formal proof of the Feit-Thompson theorem

Participants: Laurence Rideau, Laurent Théry.

The Feit-Thompson theorem, established in the beginning of the 1960s, states that every odd-order finite group is solvable. The proof of this result was initially published in an article with around 250 pages. This proof was cleaned by a team of mathematicians and re-published in the form of two books, totaling approximately the same number of pages. But these books also rested on some general knowledge about groups and various areas of algebras.

All this knowledge is now formally described in the Mathematical Components library. The proof of the theorem has been completed in September 2012. The team that achieved this result includes members of the Marelle project-team, along with members of the Typical project-team at Inria Saclay-Ile de France, members of the Microsoft Research Cambridge laboratory, and guests from other institutions.

This year, the members of the Marelle team concentrated on the following topics:

- General character theory: chapters 5 and 6 of the book by Isaacs,
- Character theory for the odd order theorem: chapters 1 to 4 of the book of the book by Peterfalvi.

More information at http://www.msr-inria.inria.fr/Projects/math-components/feit-thompson.

5.5. Native execution for the Coq system

Participants: Maxime Dénès, Benjamin Grégoire, Yves Bertot.

We have continued our work on the native execution of dependently typed terms, aiming at the integration of this work in the main branch of the Coq system.

5.6. Provably correct approximations of elementary functions

Participants: Erik Martin-Dorel, Laurence Rideau, Laurent Théry.

The elementary functions are general purpose mathematical functions that are often implemented in the hardware of modern micro-processors: exponential and trigonometric functions, and inverse functions like arctan or square-root. We participate in a nationally funded project (ANR-TaMaDi) where precise approximations of these functions and their combinations must be computed. A first approach is to use Taylor models. We implemented such an approach and proved its correctness in the Coq system. This led to the publication [9].

We are now working on applying Bernstein polynomials to the problem of approximating transcendental functions.

5.7. Geometric algebras

Participant: Laurent Théry.

We translated our library to the ssreflect setting and provided a very concise certified implementation of geometric algebras based on binary trees.

5.8. Bourbaki in Coq

Participant: José Grimm.

In previous years, we developed a formal library describing the part of the Bourbaki books on set theory, cardinals, and ordinals. The whole development now runs under Coq 8.4, ssreflect 1.4. The main contribution this year is the study of some families of numbers (Stirling numbers of the second kind, Euler numbers, Bell numbers), and their relations to cardinalities (number of partitions of a set, number of partition with p parts, number of surjections $I_n \to I_p$). We have some explicit formulas for $\sum_{i<n} i^k$ as sums of binomial coefficients.

5.9. Reasoning on polynomial expressions

Participants: José Grimm, Julianna Zsido, Yves Bertot.

Continuing previous work by Bertot, we showed that if p is a polynomial on any ordered ring, that has n positive roots, the list of its coefficients has at least n sign changes. If there is exactly one sign change, and the ring is an Archimedian field, there is a number a such that the polynomial is negative on $[0, a]$ and strictly increasing after a; thus it has at most one positive root, and there is a Cauchy sequence x_i such that $p(x_i) < 0$ but $p(x_i + c/2^n) > 0$.

The publication by Bertot, Mahboubi, and Guilhot in 2011 on Bernstein polynomials describes a procedure that works only for polynomials with simple roots. We added the proofs that describe how to obtain such polynomials, starting from arbitrary ones. In other words, we proved the following statement: *for every polynomial p, p divided by the greatest common divisor of p and its derivative has the same roots as p and all the roots are simple.*

We started working on a proof that the dichotomy process based on Bernstein polynomials is bound to terminate, concentrating on a theorem known as *the theorem of three circles*.

5.10. Higher-Order Abstract Syntax

Participant: Julianna Zsido.

With Martin Hyland from the University of Cambridge, we worked on an approach to reconcile the points of view of Fiore, Plotkin, and Turi on the one hand and Hirschowitz and Maggesi on the other hand. This approach relies on a large monad that abstracts over the two approaches.

5.11. Proofs in cryptography

Participants: Gilles Barthe [IMDEA Software Institute], Juan Manuel Crespo [IMDEA Software Institute], Benjamin Grégoire, Sylvain Heraud [Prove&Run], César Kunz [IMDEA Software Institute], Yassine Lakhnech [University of Grenoble], Pierre-Yves Strub [IMDEA Software Institute], Santiago Zanella Béguelin [IMDEA Software Institute].

We are continuing our work on providing a user-friendly tool for cryptographers who want to develop formal proofs of correctness, based on Certicrypt and SMT provers. There were invited talks at ITP, CPP, MPP, SAS, and JFLA. There was also an article in ERCIM news, whose contents is more oriented towards the open public. See also the web page http://easycrypt.gforge.inria.fr/.

As an illustrative example, we proposed a machine-checked proof of a construction of a hash function based on elliptic curves, where the correctness proof uses the Random Oracle Model. The proof is based on an extension of CertiCrypt for reasoning about approximate forms of observational equivalence and uses mathematical results from group theory and elliptic curves.

Thanks to our language-based approach to describing cryptographic constructions and our automatic approach to proving them correct, we can now explore systematically the space of possible designs. Using this approach, we have been able to explore over 1.3 million schemes, including more than 100 variants of OAEP studied in the literature and to prove the correctness of 250,000 schemes for one kind of model and 17,000 for another kind.

MEXICO Project-Team

6. New Results

6.1. Avoiding shared clocks in networks of timed automata

Networks of timed automata (NTA) are widely used to model distributed real-time systems. Quite often in the literature, the automata are allowed to share clocks, i.e. the transitions of one automaton may be guarded by a condition on the value of clocks reset by another automaton. This is a problem when one considers implementing such model in a distributed architecture, since reading clocks a priori requires communications which are not explicitly described in the model.

In [58], we focus on the following question: given an NTA $A_1 \| A_2$ where A_2 reads some clocks reset by A_1, does there exist an NTA $A'_1 \| A'_2$ without shared clocks with the same behavior as the initial NTA? For this, we allow the automata to exchange information during synchronizations only, in particular by copying the value of their neighbor's clocks.

We discuss a formalization of the problem and give a criterion using the notion of contextual timed transition system, which represents the behavior of A_2 when in parallel with A_1. Finally, we effectively build $A'_1 \| A'_2$ when it exists.

6.2. Model checking languages over infinite alphabets

In [61], we consider data words, i.e, strings where each position carries both a label from a finite alphabet and some values from an infinite domain. The latter can be used to represent an unbounded number of process identifiers so that data words are suitable to model the behavior of a concurrent program with dynamic process creation. A variety of formalisms, including logic and automata, have been studied in the literature to specify sets of data words in the context of verification. Most of them focus on the satisfiability problem of very restricted logics, as the general problem is undecidable.

Here, we consider the dual approach of restricting the domain of data words instead of pruning the logic. This allows us to tackle the model-checking problem with respect to monadic second-order (MSO) properties. As model checking is undecidable for nearly all known automata models (including the model presented in the first part of the talk), we introduce data pushdown automata (DPA). DPA come with multiple pushdown stacks (where the access to stacks is bounded by a number of phase switches) and are enriched with parameters that can be instantiated with data values. DPA can model interesting protocols like a leader election protocol with an unknown number of processes. While satisfiability for MSO logic is undecidable (even for weaker fragments such as first-order logic), we show that one can decide if all words generated by a DPA satisfy a given formula from the full MSO logic.

6.3. Construction of Hanf sentences

A classical result by Hanf from the 60s states that first-order formulas over structures of bounded degree are equivalent to boolean combinations of statements of the form: “pattern P occurs at least n times”. Hanf’s theorem has many model-theoretic applications, in particular in automata theory and database query answering.

However, until recently, no elementary construction was known. In [49], we present the first algorithm that computes a Hanf normal in elementary time. More precisely, our algorithm is triply exponential, which we also show to be optimal.

6.4. A probabilistic Kleene theorem

In [63], we establish a Kleene theorem for (Rabin) probabilistic automata over finite words. Probabilistic automata generalize deterministic finite automata and assign to a word an acceptance probability. For convenient specification of probabilistic queries, we provide probabilistic expressions with probabilistic choice, guarded choice, concatenation, and a star operator. Our expressions are closer to language-theoretic operations than previous calculi for probabilistic systems, which were rather motivated by system modeling than query answering. We prove that probabilistic expressions and probabilistic automata are expressively equivalent. Our result extends to two-way probabilistic automata with pebbles and corresponding expressions.

6.5. A temporal logic for frequency properties

In, [62], we introduce fLTL, a quantitative extension of the widely used specification language LTL that allows us to express relative frequencies by a generalization of temporal operators. This facilitates the specification of requirements such as the deadlines in a real-time system must be met in at least 95% of all cases. For our novel logic, we establish an undecidability result regarding the satisfiability problem but identify a decidable fragment which strictly increases the expressiveness of LTL by allowing, e.g., to express non-context-free properties.

6.6. Adding pebbles to weighted automata: Easy specification & efficient evaluation

In [67], we extend weighted automata and weighted rational expressions with 2-way moves and (reusable) pebbles. We show with examples from natural language modeling and quantitative model-checking that weighted expressions and automata with pebbles are more expressive and allow much more natural and intuitive specifications than classical ones. We extend Kleene-Schützenberger theorem showing that weighted expressions and automata with pebbles have the same expressive power. We focus on an efficient translation from expressions to automata. We also prove that the evaluation problem for weighted automata can be done very efficiently if the number of (reusable) pebbles is low.

6.7. MSO decidability of multi-pushdown systems via split-width

Multi-threaded programs with recursion are naturally modeled as multi-pushdown systems. The behaviors are represented as multiply nested words (MNWs), which are words enriched with additional binary relations for each stack matching a push operation with the corresponding pop operation. Any MNW can be decomposed by two basic and natural operations: shuffle of two sequences of factors and merge of consecutive factors of a sequence. We say that the split-width of an MNW is k if it admits a decomposition where the number of factors in each sequence is at most k. The MSO theory of MNWs with split-width k is decidable. In [66], we introduce two very general classes of MNWs that strictly generalize known decidable classes and prove their MSO decidability via their split-width and obtain comparable or better bounds of tree-width of known classes.

6.8. Contextual Petri nets

Contextual nets (c-nets) are an extension of Petri nets that – unlike ordinary Petri nets – faithfully models concurrent read accesses to shared resources. This is not only interesting from a semantic but also from an algorithmic point of view, as the analysis of such nets can better exploit the fact that concurrent reads are independent and concurrent. In particular, the unfolding of a contextual net may be up to exponentially smaller in certain situations.

In previous work carried out in the Mexico project, we established theoretical foundations [6] and efficient algorithms for constructing c-net unfoldings [42]. More recently, we have investigated verification techniques based on c-nets. These exploit the advantages mentioned above to obtain results more efficiently. The results have been published in the Concur 2012 conference [70]. In parallel, the development of the Cunf tool has continued, see 5.1.2 . We are currently exploring how the technique can be combined with that of merged processes [107] for further speed-ups, and its applications in diagnosis.

6.9. Expressivity and Complexity of Timed Models

In [68], we show how to reliably compute fast-growing functions with timed-arc Petri nets and data nets. This construction provides ordinal-recursive lower bounds on the complexity of the main decidable properties (safety, termination, regular simulation, etc.) of these models. Since these new lower bounds match the upper bounds that one can derive from wqo theory, they precisely characterise the computational power of these so-called "enriched" nets. In [50], we characterize the importance of resources (like counters, channels, or alphabets) when measuring the expressiveness of Well-Structured Transition Systems (WSTS). We establish, for usual classes of well partial orders, the equivalence between the existence of order reflections (non-monotonic order embeddings) and the simulations with respect to coverability languages. We show that the non-existence of order reflections can be proved by the computation of order types. This allows us to extend the current classification of WSTS, in particular solving some open problems, and to unify the existing proofs.

6.10. Concurrent Games on Infinite State Systems

In [65], we propose to study concurrent games on a new extension of Vector Addition Systems with States, where inhibition conditions are added for modeling purposes. Games are a well-suited framework to solve control problems, and concurrent semantics reflect realistic situations where the environment can always produce a move before the controller, although it is never required to do so. This is in contrast with previous works, which focused mainly on turn-based semantics. Moreover, we consider asymmetric games, where environment and controller do not have the same capabilities, although they both have restricted power. In this setting, we investigate reachability and safety objectives, which are not dual to each other anymore, and we prove that (i) reachability games are undecidable for finite targets, (ii) they are 2-EXPTIME-complete for upward-closed targets and (iii) safety games are co-NP-complete for finite, upward-closed and semi-linear targets. Moreover, for the decidable cases, we build a finite representation of the corresponding controllers.

6.11. Rare Event Analysis for Markovian Systems

Model checking real time properties on probabilistic systems requires computing transient probabilities on continuous time Markov chains. Beyond numerical analysis ability, a probabilistic framing can only be obtained using simulation. This statistical approach fails when directly applied to the estimation of very small probabilities. In [60], combining the uniformization technique and extending our previous results, we design a method which applies to continuous time Markov chains and formulas of a timed temporal logic. The corresponding algorithm has been implemented in our tool cosmos. We present experimentations on a relevant system, with drastic time reductions with respect to standard statistical model checking.

Statistical model-checking is an alternative verification technique applied on stochastic systems whose size is beyond numerical analysis ability. Given a model (most often a Markov chain) and a formula, it provides a confidence interval for the probability that the model satisfies the formula. One of the main limitations of the statistical approach is the computation time explosion triggered by the evaluation of very small probabilities. In order to solve this problem, we develop in [59] a new approach based on importance sampling and coupling. The corresponding algorithms have been implemented in our tool cosmos. We present experimentation on several relevant systems, with estimated time reductions reaching a factor of $10^{1}20$.

6.12. Conformance Relations for Labeled Event Structures

In [69], we have extended several well known conformance (ioco style) relations for sequential models, to the concurrent framework of labeled event structures. With the interleaving semantics, the relations we obtained boil down to the same relations defined for labeled transition systemss. By contrast, under the partial order semantics, the relations we obtain allow to distinguish explicitly implementations where concurrent actions are implemented concurrently, from those where they are interleaved, i.e. implemented sequentially. Therefore, these relations will be of interest when designing distributed systems, since the natural concurrency between actions that are performed in parallel by different processes can be taken into account. In particular, the fact of being unable to control or observe the order between actions taking place on different processes will not be considered as an impediment for testing.

A complete testing framework for concurrent systems has been developped. We studied what kind of systems are testable in such a framework and we have proposed sufficient conditions for obtaining a complete test suite. Finally, an algorithm to construct a test suite with such properties was proposed. These result are summarized in a paper that is being prepared for a journal submission.

MUTANT Project-Team

6. New Results

6.1. Information-Geometric Approach to Real-time Audio Change Detection

Participants: Arnaud Dessein, Arshia Cont.

We developed a generic framework for real-time change detection of audio signals using methods of information geometry. The present method is limited to generative models of audio signals based on generic exponential distribution families. The proposed system detects changes by controlling the information rate of the signal as they arrive in time. The method also addresses shortcomings of traditional approaches based on cumulative sums which assume known parameters before change. This is achieved by calculating exact generalized likelihood ratio test statistics with complete estimation of unknown parameters in respective hypothesis [9]. The interpretation of this framework within a dually flat geometry of exponential families provide tractable algorithms for online use. Results are presented for speech segmentation into different speakers and polyphonic music segmentation.

6.2. Real-time Polyphonic Music Recognition

We investigated real-time recognition of overlapping music events in two context of dictionary-based detection and real-time alignment:

6.2.1. Real-time detection of overlapping sound events using non-negative matrix factorization

Participants: Arnaud Dessein, Arshia Cont.

Non-negative matrix factorization (NMF) methods have naturally found their way since their inception to sound and music processing. This work is an extension to our previous work in [1] on Real-time Music Transcription using sparse NMF methods. We investigate the problem of real-time detection of overlapping sound events by employing NMF techniques. We consider a setup where audio streams arrive in real-time to the system and are decomposed onto a dictionary of event templates learned off-line prior to the decomposition. An important drawback of existing approaches in this context is the lack of controls on the decomposition. We propose and compare two provably convergent algorithms that address this issue, by controlling respectively the sparsity of the decomposition and the trade-off of the decomposition between the different frequency components. Sparsity regularization is considered in the framework of convex quadratic programming, while frequency compromise is introduced by employing the beta-divergence as a cost function. The two algorithms are evaluated on the multi-source detection tasks of polyphonic music transcription, drum transcription and environmental sound recognition. The obtained results in [20] show how the proposed approaches can improve detection in such applications, while maintaining low computational costs that are suitable for real-time.

A specialized version of NMF for Real-time Music Transcription is exposed in Arnaud Dessein's PhD thesis [9].

These methods will be subject to software development in 2013.

6.2.2. Robust Real-time Polyphonic Audio-to-Score Alignment

Participant: Arshia Cont.

The *Antescofo* system is polyphonic since 2009 but its use in highly polyphonic and noisy concert environments have been challenging. To overcome this, we have studied more robust inference mechanisms. As a results, the previous inference mechanism based on maximum a posteriori of Viterbi Forward variables in mixed semi-Markov and Markov chains in [2] were abandoned in favor of a more robust method based on *importance resampling* on state-space models and smoothing of variable-order hybrid chains. This has led to robust real-time alignment and the employment of the system in various Piano performances in 2012. Further extensions are currently under study.

6.3. Real-time Multi-object Detection for Music Signals

Participants: Philippe Cuvillier [Master 2 ATIAM], Arshia Cont.

Multiple-object detection and tracking has been widely used in applications such as missile tracking and radar and has given birth to several formalisms such as Random Finite Sets [33]. Such formalisms can be seen as extensions to existing probabilistic inference mechanisms with explicit birth and death stochastic mechanisms for multiple source tracking.

In this work we aim at studying such formalisms in the case of real-time music signal processing. The idea is to track multiple sources (instruments, audio flows) from one source of observation. This approach can be beneficial to two main applications in real-time music listening:

- Extension of existing audio-to-score [2] or audio-to-audio alignment [7] mechanisms (currently based on one source) to multiple objects can address the following short-comings of existing approaches: explicit consideration for asynchrony of parallel sources; robustness to uncertainties on one or more voices.
- Studying the classical *Partial Tracking* applications in audio processing within the RFS context can lead to better results in low-level sinusoidal partial tracking of sounds.

Early studies of such formalisms are exposed in [25]. Concrete applications will be exposed in 2013.

6.4. Antescofo Language Extensions and Performance Fault-Tolerance

We have improved the *Antescofo* framework widely used for mixed instrumental and live electronic computer music. The new framework paves the way for future language extensions and paves the way for future research regarding performance fault-tolerance, synchronization mechanisms and formal verifications.

6.4.1. Antescofo Language Extensions

Participants: José Echeveste, Jean-Louis Giavitto, Florent Jacquemard, Arshia Cont.

To further extend the *Antescofo* language, the system has been formally modeled as a network of parametric timed automata in [29]. The model obtained provides operational semantics for the input scores, in particular the interaction between the instrumental and electronic parts and the timing and error handling strategies mentioned below. This approach would enable better authoring of time and interaction during programing/composing, permits to use state of the art software verification tools for the static analysis of Antescofo scores and also provides means to address critical aspects of musical performances in real-time.

In parallel, a new grammar for the score language and a new architecture have been designed for *Antescofo*, taking into account new demands from the community such as addition of timed variables in the language, dynamic time processes, time-conditional constructs, and more.

6.4.2. Performance Fault-Tolerance and Synchronization Mechanisms

Participants: José Echeveste, Jean-Louis Giavitto, Arshia Cont.

We formalized the timing strategies for musical events taking into account the variability of environment signals (musicians) and their effect on computer events programmed in *Antescofo*. The result of this work is presented in [15], where new block attributes in the language determine expected behavior in case of environment changes in real-time (errors, timing discrepancies, etc.). These additions have been implemented in the current version of the system and are widely used by the user community.

6.5. Temporal Analysis and Verification of Interactive Music Scores

Participants: Léa Fanchon [Master 2 École Centrale], Florent Jacquemard.

Léa Fanchon's Masters thesis, under the supervision of Florent Jacquemard, [26] presents an analysis module that complements the real-time score authoring and performance in *Antescofo*, with the aim of exploring possible behavior of authored programs with respect to possible deviations in human musician performance. This work employs formal methods for temporal automata networks using linear constraint inference techniques commonly in use for task scheduling and circuit verifications.

Obtained results pave the way for future works in formal verification of interactive multimedia applications, being one of the first of its kind in computer music literature, and provides the following input to programmers and artists using *Antescofo*:

- Evaluation of robustness of the program with respect to the environment's (musician's performance) temporal variations,
- Feedback to programmers/artists on critical synchronization points for better programming.

An article describing this work is currently in preparation for a submission to a computer music conference.

6.6. Formal study of Antescofo as a Reactive System

Participants: Guillaume Baudart [Master 2 ATIAM], Florent Jacquemard, Marc Pouzet [ENS], Jean-Louis Giavitto, Arshia Cont.

An *Antescofo* score/program can be considered as a specification of a reactive system through its coupling of a machine listening with a real-time synchronous language. In his master thesis under the supervision of Florent Jacquemard and Marc Pouzet (team Parkas), Guillaume Baudart has has studied the links between the reactive system of *Antescofo* and existing synchronous languages such as *Lucid Synchrone* [36] and *Reactive ML* [34]. The reactive engine of a preliminary version of *Antescofo* was developed in both languages and their structures were compared.

This study reveals the particularities of musical applications of reactive systems specific to *Antescofo* (see [24]). *Reactive ML* allows dynamic constructions but real-time performance can not be guaranteed especially when the machine listening is combined with the reactive system. On the contrary, *Lucid Synchrone* does not easily allow dynamic process creation. Each language specificity leads to strong considerations in the program/score structure for the artists. This work will be continued in 2013 to further strengthen ties between the reactive aspects of *Antescofo* and that of synchronous languages.

6.7. Tree Structured Presentation of Symbolic Temporal Data

Participants: Florent Jacquemard, Michael Rusinowitch [Project-team Cassis], Luc Segoufin [Project-team Dahu].

In traditional music notation, in particular in the languages used for the notation of mixed music such as Antescofo DSL, the durations are not expressed by numerical quantities but by symbols representing successive subdivisions of a reference time value (the beat). For this reason, trees data structures are commonly used for the symbolic representation of rhythms in computer aided composition softwares such as OpenMusic (developed at Ircam). It is therefore worth studying the applications in rhythm notation of existing formalisms for recognizing, querying, transforming and learning sets of tree structured data.

In 2012 we have studied several classes of tree recognizers which could be of interest in this context. First, with Michael Rusinowitch we have proposed in [16] a novel class of automata computing on unranked trees, which are context free in two dimensions: in the the sequence of successors of a node and also along paths. Second, we studied with Luc Segoufin [21] automata and logics computing on data trees and their relationship. Data trees are unranked ordered trees where each node carries a label from a finite alphabet and a datum from some infinite domain.

PAREO Project-Team

6. New Results

6.1. Model transformation

Participants: Jean-Christophe Bach, Pierre-Etienne Moreau.

In [10], we have proposed a general method to transform high level models by using *Tom* strategies. High-level models we consider are *EMF-ECore* models that we represent by terms whose mappings have been generated by the *Tom-EMF* tool. The proposed method consists in decomposing a complex transformation into many elementary transformations (*definitions*) encoded by *Tom* strategies. These *definitions* are applied on a source model without any consideration of execution order. Therefore, we proposed a mechanism to address the problem of dependency between elementary transformations without introducing any scheduling between rewriting rules. This mechanism relies on the use of temporary elements which play the roles of the target elements until the last part of the transformation : the *Resolve* phase. The goal of this phase is to find and replace all temporary elements by real target ones, and therefore to reconnect all partial target models obtained during elementary transformations to build the resulting model.

In [11], [15], we presented a first proposal of a high-level transformation language included in *Tom* which implements the aforementioned general method. We used this language to implement an avionic case study — AADL2Fiacre — which was proposed by Airbus for the *quarteFt* project.

6.2. Improvements of theoretical foundations

6.2.1. Termination under strategies

Participants: Horatiu Cirstea, Pierre-Etienne Moreau.

Several approaches for proving the confluence and the termination of term rewriting systems have been proposed [16] and the corresponding techniques have been implemented in tools like Aprove [23] and TTT2 [32]. On the other hand, there are relatively few works on the study of these properties in the context of strategic rewriting and the corresponding results were generally obtained for some specific strategies and not within a generic framework. It would thus be interesting to reformulate these notions in the general formalism we have previously proposed [21] and to establish confluence and termination conditions similar to the ones used in standard rewriting.

We have first focused on the termination property and we targeted the rewriting strategies of the *Tom* language. We propose a direct approach which consists in translating *Tom* strategies into a rewriting system which is not guided by a given evaluation strategy and we show that our systematic transformation preserves the termination. This allowed us to take advantage of the termination proof techniques available for standard rewriting and in particular to use existing termination tools (such as Aprove and TTT2) to prove the termination of strategic rewriting systems. The efficiency and scalability of these latter tool has a direct impact on the performances of our approach especially for complex strategies for which an important number of rewrite rules could be generated. We have nevertheless proposed a meta-level implementation of the automatic transformation which improves significantly the performances of the approach.

6.2.2. Automatizing the certification of induction proofs

Participant: Sorin Stratulat.

Largely adopted by proof assistants, the conventional induction methods based on explicit induction schemas are non-reductive and local, at schema level. On the other hand, the implicit induction methods used by automated theorem provers allow for lazy and mutual induction reasoning. In collaboration with Amira Henaien [13], we devised a new tactic for the Coq proof assistant able to perform automatically implicit induction reasoning. By using an automatic black-box approach, conjectures intended to be manually proved by the certifying proof environment that integrates Coq are proved instead by the `Spike` implicit induction theorem prover. The resulting proofs are translated afterwards into certified Coq scripts.

As a case study, conjectures involved in the validation of a non-trivial application [35] have been successfully and directly certified by Coq using the `Spike` tactic. The proofs of more than 60% of them have been performed completely automatically, i.e., the Coq user does not need to provide any argument to the tactic. On the other hand, its application is limited to Coq specifications transformable into conditional specifications whose axioms can be oriented into rewrite rules.

6.2.3. Cyclic proofs by induction methods

Participant: Sorin Stratulat.

In a first-order setting, two different 'proof by induction' methods are distinguished: the conventional induction, based on explicit induction schemas, and the implicit induction, based on reductive procedures. In [14], we proposed a new cycle-based induction method that keeps their best features, i.e., performs local and non-reductive reasoning, and naturally fits for mutual and lazy induction. The heart of the method is a proof strategy that identifies in the proof script the subset of formulas contributing to validate the application of induction hypotheses. The conventional and implicit induction are particular cases of our method.

6.3. Integration of formal methods in programming languages

6.3.1. Multi-focus strategies

Participants: Jean-Christophe Bach, Christophe Calvès, Horatiu Cirstea, Pierre-Etienne Moreau.

Like most rewriting engines, *Tom* patterns combined with traversal strategies, gives the possibility to match and rewrite at any position in a given term. We have extended this classical approach with multi-focus strategies which enable us to match and rewrite several positions simultaneously. More precisely, the action performed at a given position can depend on the other positions involved in the corresponding strategy. This extension is particularly well-suited for programming-language semantics specification, semantics which usually require gathering several subterms (code, memory, input/output channels, ...) to perform one action.

The multi-focus library is a conservative extension of *Tom* standard strategies and provides combinators to handle multi-position traversal, matching and rewriting. Compared to the original *Tom* strategy library, the multi-focus version provides global backtracking. The library is available at http://gforge.inria.fr/projects/tom.

6.3.2. Formal islands grammars parsing

Participants: Jean-Christophe Bach, Pierre-Etienne Moreau.

Extending a language by embedding within it another language presents significant parsing challenges, especially if the embedding is recursive. The composite grammar is likely to be nondeterministic as a result of tokens that are valid in both the host and the embedded language. In [9], we examined the challenges of embedding the *Tom* language into a variety of general-purpose high level languages. The current parser of *Tom* is complex and difficult to maintain. In this paper, we described how *Tom* can be parsed using island grammars implemented with the Generalised LL (*GLL*) parsing algorithm. The grammar is, as might be expected, ambiguous. Extracting the correct derivation relies on a disambiguation strategy which is based on pattern matching within the parse forest. We described different classes of ambiguity and proposed patterns to solve them.

6.4. Security policies specification and analysis

Participants: Horatiu Cirstea, Hélène Kirchner, Pierre-Etienne Moreau.

Access control policies, a particular case of security policies should guarantee that information can be accessed only by authorized users and thus prevent all information leakage. We proposed [12] a framework where the security policies and the systems they are applied on are specified separately but using a common formalism. This separation allows not only some analysis of the policy independently of the target system but also the application of a given policy on different systems. In this framework, we propose a method to check properties like confidentiality, integrity or confinement over secure systems based on different policy specifications.

PARKAS Project-Team

6. New Results

6.1. Reactive Programming

Participants: Mehdi Dogguy, Louis Mandel, Cédric Pasteur, Marc Pouzet.

ReactiveML is an extension of OCaml with synchronous concurrency, based on synchronous parallel composition and broadcast of signals. The goal is to provide a general model of deterministic concurrency inside a general purpose functional language to program reactive systems. It is particularly suited to program discrete simulations, for instance of sensor networks.

One of the current focus of the research is being able to simulate huge systems, composed of millions of agents, by extending the current purely sequential implementation in order to be able to take advantage of multi-core and distributed architectures. This goal has led to the introduction of a new programming construct, *reactive domain*, which allows to define local time scales. These domains help for the distribution of the code but also increase the expressiveness of the language. In particular, it allows to do time refinement. A paper on this new construct and the related static analysis has been submitted. We have implemented a new runtime for ReactiveML, that uses the MPI (Message Passing Interface) library to run programs on multi-core and distributed architectures.

We have also investigated new static analyses for the language. Following the work of PhD thesis of Mehdi Dogguy, we have studied a new analysis which adds usages on signals to be able to ensure one to one communications. We have also studied a new reactivity analysis which ensures that a process can not prevent the other ones to from executing. This analysis will be published in [10].

6.2. n-Synchronous Languages

Participants: Louis Mandel [contact], Marc Pouzet, Albert Cohen, Adrien Guatto.

The n-synchronous model introduced a way to compose streams which have *almost the same clock* and can be synchronized through the use of a finite buffer.

We have designed the language Lucy-n to program in this model of computation [40]. This language is similar to the first order synchronous data-flow language Lustre in which a buffer operator is added. A dedicated type system allows to check that programs can be executed in bounded memory and to compute sufficient buffer sizes. Technically it is done through the introduction of a subtyping constraint at each bufferization point.

- In collaboration with F. Plateau (Prove&Run), we developed a new resolution constraint algorithm for the clocking of Lucy-n programs [8]. Even if the new algorithm is less efficient that the one using abstraction, it has the advantage to be more precise and thus to accept more programs. It is useful for example for the static scheduling of Latency Insensitive Designs [41].
- We worked on an extension of the synchronous model with integer clocks. This extension allows to produce and consume several values at each activation. It has large implication on the semantics, clock typing, causality and code generation of the language.
- We have continue the work on the code generation. In particular, we have been designing a new intermediate representation that allows to deal with integer clocks.

6.3. Strong normal form for large integers, boolean functions and finite automata

Participant: Jean Vuillemin.

Jean Vuillemin's recent work focusses on finding Strong Normal Form for large Integers, Boolean functions and finite Automata, with applications to circuits and software.

- [16] is the latest version of JV's course notes at ENS "De l'algorithme au circuit".
- [9] shows that the ordered dimension of a Boolean function is a lower bound on the size of most known ordered Decision Diagrams, and that ordered decision diagrams can be efficiently constructed an operated upon.
- [6] shows an approach to circuit protection against side-channel attacks based on a statistical analysis of power traces derived from actual measures of the circuit in operation.

6.4. A theory of safe optimisations in the C11/C++11 memory model and applications to compiler testing

Participants: Francesco Zappa Nardelli [contact], Robin Morisset, Pankaj Pawan.

Compilers sometimes generate correct sequential code but break the concurrency memory model of the programming language: these subtle compiler bugs are observable only when the miscompiled functions interact with concurrent contexts, making them particularly hard to detect. In this work we design a strategy to reduce the hard problem of hunting concurrency compiler bugs to differential testing of sequential code and build a tool that puts this strategy to work. Our first contribution is a theory of sound optimisations in the C11/C++11 memory model, covering most of the optimisations we have observed in real compilers and validating the claim that common compiler optisations are sound in the C11/C++11 memory model. Our second contribution is to show how, building on this theory, concurrency compiler bugs can be identified by comparing the memory trace of compiled code against a reference memory trace for the source code. Our tool identified several mistaken write introductions and other unexpected behaviours in the latest release of the gcc compiler.

A paper on this work has been submitted to an international conference [15].

6.5. A verified compiler for relaxed-memory concurrency

Participant: Francesco Zappa Nardelli [contact].

We studied the semantic design and verified compilation of a C-like programming language for concurrent shared-memory computation above x86 multiprocessors. The design of such a language is made surprisingly subtle by several factors: the relaxed-memory behaviour of the hardware, the effects of compiler optimisation on concurrent code, the need to support high-performance concurrent algorithms, and the desire for a reasonably simple programming model. In turn, this complexity makes verified (or verifying) compilation both essential and challenging. This project started in 2010, and in 2012 we submitted a journal version, describing the correctness proof of all the phases of our CompCertTSO compiler (including experimental fence eliminations). This has been accepted for publication in Journal of the ACM [3].

In collaboration with Jaroslav Sevcik (U. Cambridge), Viktor Vafeiadis (MPI-SWS), Suresh Jagannathan (Purdue U.), Peter Sewell (U. Cambridge).

6.6. Compiling C/C++ concurrency from C++11 to POWER

Participant: Francesco Zappa Nardelli [contact].

The upcoming C and C++ revised standards add concurrency to the languages, for the first time, in the form of a subtle relaxed memory model (the C++11 model). This aims to permit compiler optimisation and to accommodate the differing relaxed-memory behaviours of mainstream multiprocessors, combining simple semantics for most code with high-performance low-level atomics for concurrency libraries.

We studied the the correctness of two proposed compilation schemes for the C++11 load and store concurrency primitives to Power assembly, having noted that an earlier proposal was flawed. (The main ideas apply also to ARM, which has a similar relaxed memory architecture.)

This should inform the ongoing development of production compilers for C++11 and C1x, clarifies what properties of the machine architecture are required, and builds confidence in the C++11 and Power semantics.

A paper describing this work will appear in POPL 2012 [5].

In collaboration with Kayvan Memarian (previously student in the Moscova EPI, currently at U. Cambridge).

6.7. Compilation techniques for synchronous languages

Participants: Marc Pouzet [contact], Adrien Guatto, Léonard Gérard, Cédric Pasteur.

- The generation of efficient sequential code for synchronous data-flow languages raises two intertwined issues: control and memory optimization. While the former has been extensively studied, for instance in the compilation of Lustre and SIGNAL, the latter has been only addressed in a restricted manner. Yet, memory optimization becomes a pressing issue when arrays are added to such languages, for example, SCADE 6 [8]. We have proposed a two-levels solution to the memory optimization problem. It combines a compile-time optimization algorithm, reminiscent of register allocation, paired with language annotations on the source given by the designer. Annotations express in-place modifications and control where allocation is performed. Moreover, they allow external functions performing in-place modifications to be imported safely. Soundness of annotations is guaranteed by a semilinear type system and additional scheduling constraints. A key feature is that annotations for well-typed programs do not change the semantics of the language: removing them may lead to a less efficient code but with the very same semantics.

  The method has been implemented in HEPTAGON, the compiler developed in the team of a Lustre-like synchronous language extended with hierarchical automata and arrays. Experiments show that the proposed approach removes most of the unnecessary array copies, resulting in faster code that uses less memory. This work has been presented at the *ACM Intern. Conf. on Languages, Compilers and Tools for Embedded Systems (LCTES'12)* in June 2012 and it has received the *Best paper award*.

6.8. Generation of Parallel Code from Synchronous Programs

Participants: Albert Cohen [contact], Léonard Gérard, Adrien Guatto, Nhat Minh Le, Marc Pouzet.

- Efficiently distributing synchronous programs is a challenging and long-standing subject. This paper introduces the use of futures in a Lustre-like language, giving the programmer control over the expression of parallelism. In the synchronous model where computations are considered instantaneous, futures increase expressiveness by decoupling the beginning from the end of a computation. Through a number of examples, we show how to desynchronize long computations and implement parallel patterns such as fork-join, pipelining and data parallelism. The proposed extension preserves the main static properties of the base language, including static resource bounds and the absence of deadlock, livelock and races. Moreover, we prove that adding or removing futures preserves the underlying synchronous semantics.

  This work has been presented at the *ACM Intern. Conf. on Embedded Software (EMSOFT 2012)*, in October 2012 and it received the *Best paper award*.

  Further work along these lines is taking place, to generate code for a variety of low-overhead execution models, to cope with real-time constraints, and to formalize and prove the correctness of the underlying concurrent data structures. On the latter point, a paper has been accepted at the ACM Conf. PPoPP 2013.

6.9. Semantics and Implementation of Hybrid System Modelers

Participants: Marc Pouzet [contact], Timothy Bourke.

[8] http://www.esterel-technologies.com/products/scade-suite/

Zélus is a new programming language for modeling systems that mix discrete logical time and continuous time behaviors. From a user's perspective, its main originality is to extend an existing -like synchronous language with Ordinary Differential Equations (ODEs). The extension is conservative: any synchronous program expressed as data-flow equations and hierarchical automata can be composed arbitrarily with ODEs in the same source code. A dedicated type system and causality analysis ensure that all discrete changes are aligned with zero-crossing events so that no side effects or discontinuities occur during integration. Programs are statically scheduled and translated into sequential code which, by construction, runs in bounded time and space. Compilation is effected by source-to-source translation into a small synchronous subset which is processed by a standard synchronous compiler architecture. The resulting code is paired with an off-the-shelf numeric solver.

This experiment show that it is possible to build a modeler for explicit hybrid systems à la Simulink/Stateflow on top of an existing synchronous language, using it both as a semantic basis and as a target for code generation. In parallel with the software development done during the year, we investigate, in collaboration with Albert Benveniste, Benoit Caillaud (Inria Rennes) and Dassault-Systèmes the treatment of Differential Algebraic Equations (DAEs), in explicit or semi-explicit form.

This work will be presented at the *ACM Intern. Conference on Hybrid Systems: Computation and Control (HSCC 2013)* in April 2013.

PARSIFAL Project-Team

6. New Results

6.1. Recovering Proof Structures in the Sequent Calculus

Participants: Kaustuv Chaudhuri, Stefan Hetzl, Dale Miller.

The *sequent calculus* is often criticized as a proof syntax because it contains a lot of noise. It records the precise minute sequence of operations that was used to construct a proof, even when the order of some proof steps in the sequence is irrelevant and when some of the steps are unnecessary or involve detours. These features lead to several technical problems: for example, cut-elimination in the classical sequent calculus LK, as originally developed by Gentzen, is not confluent, and hence proof composition in LK is not associative. Many people choose to discard the sequent calculus when attempting to design a better proof syntax with the desired properties.

In recent years, there has been a project at Parsifal to recover some of these alternative proof syntaxes by imposing a certain abstraction over sequent proofs. The earliest example of this was in [37], where we showed a class of sequent proofs that were isomorphic to proof nets for multiplicative linear logic. In 2012, we were able to obtain a similar result for first-order classical logic, wherein we defined a class of sequent proofs that are isomorphic to expansion trees, a generalization of Herbrand disjunctions that is in some sense a minimalistic notion of proof for classical logic. This result was published at the CSL 2012 conference [22] and a journal version is in preparation.

Our technique for recovering these dramatically different proof structures directly in the sequent calculus involves the use of *maximal multi-focusing* which gives a syntactic characterization of those sequent proofs that: (1) have a “don’t care” ordering of proof steps where the order does not matter, and (2) groups larger logical steps, called *actions*, into a maximally parallel form where only important orderings of actions are recorded. This technique was pioneered at Parsifal, and we have barely scratched the surface of its applications.

6.2. Compact Proof Certificates By Bounded Contractions

Participant: Kaustuv Chaudhuri.

An important engineering question in the ProofCert project is that of communicating, manipulating, and storing formal proof certificates. A fully detailed proof certificate, especially one generated by proof search, can be very large. Using such proofs would require a high bandwidth interface between the proof producer and consumer, which limits the scalability of the *ensemble of proving systems* approach. It is therefore natural to ask if there are more compact formats for proof certificates. The ideal format would have a tunable level of detail, so that the size of the certificates can be tailored to the application domain.

Suppose the proof consumer is equipped with some proof search capabilities. What then needs to be transmitted to the consumer to guarantee that it can check a proof within desired complexity bounds? It turns out that there is a systematic and general answer to this problem: use *focusing* and record only the “decision” rules of focusing in the proof certificate. From a high level perspective, this answer is equivalent to designing a proof system where the contraction rules are carefully bounded.

A proposal along these lines was published at the CPP 2012 conference [21]. In fact, this paper solves a harder than necessary problem by building proof certificates for linear logic, where unconstrained proof search has very high complexity even in the propositional fragment. The proposed solution is a spectrum of certificates that trades off the size of the certificate for the complexity of checking the certificate. At one end we have a very compact certificate that essentially amounts to a maximum depth of the proof, but reconstructing a proof with only a depth bound tends to be infeasible as the search space grows super-exponentially with the depth. Certificates at other end of the spectrum contain information about all the contractions in the proof; these certificates can be checked deterministically, in time proportional to the size of the certificate. Moreover, there is a simple abstraction mechanism between different levels of detail in this spectrum that allows for a *proof elaborator* to alter the level of detail in the certificate.

6.3. A Two-level Approach to Reasoning about Computation

Participant: Dale Miller.

In a paper that appeared in the J. of Automated Reasoning, Gacek, Miller, and Nadathur [12] described the foundations and architecture of a new interactive theorem prover capable of reasoning with rich collections of inductive and coinductive relations. This prover, called Abella, also contains the "generic" quantifier ∇ that provides a direct and elegant treatment of term-level binding.

A novel aspect of Abella is that it can define provability in various simple logics and can also reason about provability in such logics. The current system includes a *specification logic* that is a (restricted) intuitionistic logic programming language (a sublanguage of λProlog). The main logic of Abella is then the second logic, called the *reasoning logic*, and it is capable of reasoning about provability in the specification language.

This approach to reasoning about computation has interesting applications. For example, the reasoning logic is aware of the fact that the cut and substitution rules can be eliminated in the specification logic. As a consequence, the notoriously difficult "substitution lemmas" that occur repeated in the study of operational semantics are proved essentially for free (that is, they are an immediate consequence of cut-elimination).

In [17], Accattoli showed that when one reasons about the *untyped*λ-calculus, the specification logic is often not needed. In particular, Accattoli reinterpreted the formalization by G. Huet of the meta-theory of λ-calculus residuals in Abella and showed that the resulting meta-theory had a much more elegant and natural specification than the one presented early by Huet in Coq. While the use of two-levels of logic was not important for this particular (untyped) example, other aspects of Abella—relation specifications, ∇-quantification, and strong induction principles—were critical for improving the expressivity of this prover.

6.4. A Non-local Method for Robustness Analysis of Floating Point Programs

Participants: Dale Miller, Ivan Gazeau.

Programs that must deal with floating point programs and their associate errors can have erratic behavior. In particular, a program that yields outputs that depend continuously on their inputs (in an idealized arithmetic setting) can behave non-continuously when using floating point arithmetic. There are few tools for reasoning about program correctness in a setting that allows for such discontinuous operators.

In [23], Gazeau, Miller, and Palamidessi provide an approach to reason about some programs that are not continuous. In that paper, they introduce the notion of "robustness", which intuitively means that if the input to the program changes less than a fixed small amount then the output changes only slightly. This notion is useful in the analysis of rounding error for floating point programs because it helps to establish bounds on output errors introduced by both measurement errors and by floating point computation. Compositional methods often do not work since key constructs—like the conditional and the while-loop—are not robust. The authors proposed a method for proving the robustness of a while-loop. This method is non-local in the sense that instead of breaking the analysis down to single lines of code, it checks certain global properties of its structure. This paper shows that both the CORDIC computation of the cosine and Dijkstra's shortest path algorithm are robust.

6.5. Herbrand Confluence

Participants: Stefan Hetzl, Lutz Straßburger.

It is well-known that cut-elimination in the sequent calculus for classical first-order logic is in its most general form, is neither confluent nor strongly normalizing. But if one takes a coarser (and mathematically more realistic) look at cut-free proofs, one can analyze which witnesses they choose for which quantifiers, or in other words: one can only consider the Herbrand-disjunction of a cut-free proof. This yields a surprising confluence result for a natural class of proofs: all (possibly infinitely many) normal forms of the non-erasing cut reduction lead to the same Herbrand-disjunction. This result has been presented at CSL 2012 [25].

6.6. Semi-Star-Autonomous Categories

Participants: Willem Heijltjes, Lutz Straßburger.

A curious aspect of Girard's proof nets for multiplicative linear logic without units is that, despite being a canonical representation of proof, their categorical semantics is not obvious—this in contrast to the situation *with* units, where star-autonomous categories form a natural semantics, but no canonical proof nets are known.

In the middle of the past decade several proposals for a categorical semantics of proof nets, a notion of *semi-star-autonomous* categories, were investigated: by Robin Houston and Dominic Hughes, by Kosta Došen, and by François Lamarche and Lutz Straßburger.

The present effort by Willem Heijltjes and Lutz Straßburger completes the notion in such a way that proof nets constitute the *free* semi-star-autonomous category.

6.7. Foundations and applications of explicit substitutions

Participant: Beniamino Accattoli.

Starting from the study of Linear Logic proof nets, a new approach to explicit substitutions for ł-calculus has recently been introduced by Accattoli and D. Kesner [31]. This approach has been systematically explored by Accattoli and his co-authors.

The rewriting theory of these new explicit substitutions *at a distance* has been studied in [11] and [16]. In [11] Accattoli and Kesner study the preservation of λ-calculus strong normalization (PSN) when explicit substitutions are extended with permutative axioms allowing to swap constructors in the term, generalizing considerably the already difficult case of PSN with composition of substitutions. In [16] Accattoli developed an abstract technique for proving factorizations theorems for generic explicit substitution calculi. The factorization theorem for λ-calculus says that any reduction can be re-organized as an *head* reduction followed by a non-head reduction.

In [16] it is shown how to prove this theorem in an uniform way for many explicit subsitutions calculi. The technique emerged as a generalization of the proofs for explicit substitutions at a distance, which are simpler than usual explicit substitutions and thus lead to cleaner and more compact arguments, easier to generalize.

Applications of explicit substitutions at a distance have been studied in [19], [18], [20]. In [19] Accattoli and Dal Lago show that the length of the head reduction in calculi at a distance is a measure of time complexity. More precisely, they show that such a quantity is polynomially related (in both directions) to the cost of evaluating with Turing Machines. This result is an important step forward towards the solution of the long-standing open problem of finding a time cost model for ł-calculus.

In [20] Accattoli and Paolini apply substitutions at a distance in a call-by-value setting. They show that in this new framework there is a natural characterization of *solvability*, an important notion related to denotational semantics and the representation of partial recursive functions. In [26] (a work presented to a workshop and currently submitted to the post-proceedings of the workshop) Accattoli shows the tight relations between the framework in [20] and linear logic proof nets, providing a new characterization of the proof nets representing the call-by-value λ-calculus.

Finally, in [18] Accattoli and Kesner introduce a calculus generalizing many different extensions of λ-calculus with permutations, appeared in various contexts (studies about call-by-value, postponing of reductions, monadic languages, etc) and prove confluence and preservation of strong normalization, exploiting and extending their own results in [11].

6.8. Sequent Calculus with Calls to a Decision Procedure

Participants: Mahfuza Farooque, Stéphane Lengrand.

In the PSI project, we have designed a version of the focussed sequent calculus (for first-order classical logic) that can call external decision procedures. Since the last Activity Report, we have finished proving the essential meta-theory for it: soundness, invertibility of asynchronous rules, cut-elimination, the fact that polarities do not affect provability but only the shape of proofs, and finally completeness.

The first properties are the object of [27], while the latter ones have been obtained later in 2012.

A side-product of this meta-theory is a technical device that could be used to encode other techniques from automated reasoning like *connection tableaux*.

Secondly, we have encoded the SMT-solving algorithm DPLL(T) as the incremental construction of proof-trees in that sequent calculus [29], [28]. A very interesting aspect of the encodings is that the basic rules of DPLL(T) makes use of cuts on atoms in sequent calculus, while the advanced jrules (e.g. backjumping) makes use of general cuts. This sheds a new light on the computational speed-ups that those advanced rules provide.

We have done the encoding for two distinct presentations of DPLL(T) in the literature, and we have formalised the connection between those two descriptions [29].

6.9. Martin-Löf Identity Type in the Category of Small Categories

Participant: François Lamarche.

For the last five or six years there has been a surge of interest in finding models for the identity type in Martin-Löf type theory, and it has been clear for some time that there was a tight connection with path objects in abstract homotopy theory. A lot of proposals have been made, but there are very few semantics that fit the necessary requirements of having dependent products and also an identity type which is fully stable under substitution. The most famous model of the sort is the one proposed by Voevodsky, in his Univalent Foundations project, which uses for base category the category of simplicial sets and models dependent types by the means of Kan Fibrations. In [13] François Lamarche proposes another such model, where the base category is the category of small categories, and dependent types are modelled with Grothendieck bifibrations (maps between categories that are Grothendieck fibrations as well as their duals between the opposite categories). The full requirements of modelling Martin-Löf type theory are met. Calculations show that the model shows some amount of degeneracy "in dimensions above 2" for the associativity of equality (which should not be strict in any dimension), which is a great improvement over the models on strict groupoids and strict ω-groupoids. The construction that models the identity type is a concrete path functor for categories. It is showing itself to be very useful in homotopy theory.

PI.R2 Project-Team

6. New Results

6.1. Proof-theoretical and effectful investigations

Participants: Federico Aschieri, Pierre Boutillier, Pierre-Louis Curien, Hugo Herbelin, Guillaume Munch-Maccagnoni, Pierre-Marie Pédrot, Alexis Saurin, Arnaud Spiwack.

6.1.1. Sequent calculus and computational duality

System L syntax. Pierre-Louis Curien studied in some detail the differences (and translations) between variants of "system L" syntax for polarised classical logic (developed by Guillaume Munch-Maccagnoni and himself):

- weakly focalised systems (where negatives can be worked on at any moment in a proof) versus focalised systems (where negative and positive phases alternate strictly), versus strongly focalised systems (where furthermore negative phases have to decompose negatives completely);
- systems where changes of polarity are implicit (like in Girard's LC) versus systems where they are explicitly marked using shift operators. These shift operators are formally adjoint, and as a matter of fact a suitable intuitionistic fragment of system L corresponds exactly to Levi's CBPV;
- systems with stoup (which retain only proofs that follow the focalisation discipline) versus (still focalised) systems without stoup (where the focalisation is forced by the dynamics of reduction);
- one-sided systems (with an implicit negation given by De Morgan duality) versus two-sided systems (allowing for explicit negation, and for distinguishing the left/right and the positive/negative dualities).

Pierre-Louis Curien is also currently studying a polarised version of a notion of general connective suggested earlier by Hugo Herbelin (unpublished work), and the composition structure of these connectives (in the spirit of operads).

Categorical semantics. Guillaume Munch-Maccagnoni investigated a notion of "direct style" for adjunction models, inspired by his work on polarisation in the "L" system, in the lineage of Führmann's [47] direct-style characterisation of monadic models. (It is part of joint work with Marcelo Fiore and Pierre-Louis Curien.)

Polarised Peano arithmetic. Guillaume Munch-Maccagnoni investigates the computational contents of polarised classical logic in arithmetic and in natural deduction. This allows him to compare the constructivisation of the principle $\neg\forall \Rightarrow \exists\neg$ based on classical realisability (Krivine) and the one based on delimited control (via "double negation shift"); both of which seem to be simplified by a better understanding of the "formulae-as-types" paradigm for a negation which is involutive in a strong sense.

Guillaume Munch-Maccagnoni investigates how a notion of classical realisability structure (inspired by Krivine's) can be used to prove properties of type systems which are usually regarded as syntactic.

Classical call-by-need and the duality of computation. In 2011, Zena Ariola, Hugo Herbelin and Alexis Saurin characterised the semantics of call-by-need calculus with control in the framework of *the duality of computation*. The same set of authors extended with Paul Downen and Keiko Nakata worked on abstract machines and continuation-passing-style semantics for call-by-need with control, resulting into a paper presented at FLOPS 2012 [20].

Further work has been done by Zena Ariola, Hugo Herbelin, Luís Pinto, Keiko Nakata and José Espírito Santo on typing the continuation-passing-style of call-by-need calculus, opening the way to a proof of normalisation of simply-typed call-by-need with control, and from there to a proof of consistency of classical arithmetic with dependent choice.

Zena Ariola also investigated how to formulate a parametric theory which encompasses call-by-value, call-by-name and call-by-need. Each theory is obtained by giving the appropriate definition of what is a value and a co-value. The theory also includes so called lifting axioms which allow one to relax the syntactic restrictions previously imposed on the call-by-value, call-by-name and call-by-need calculi. The theory also allows to include the η-rules which before were causing confluence to fail. The approach can be applied to natural deduction and this allows to express different embeddings of natural deduction into sequent calculus directly in the theory. The advantage of the new formalisation is that analogously to natural deduction, one can experiment with different strategies starting from the same term. Moreover, the theory is well-suited for continuation passing style transformation and, in particular, it leads to a different and simpler formalisation of classical call-by-need, its abstract machine and continuation passing style.

6.1.2. Dependent monads

Pierre-Marie Pédrot generalised the notion of monad in order to be able to use it in a dependent framework. This new structure allows to write effects in a pure functional language, such as Coq, through a monadic encoding.

This way, the whole monadic apparatus can be lifted to dependent programs, as well as proofs.

6.1.3. Linear dependent types

Arnaud Spiwack continued his investigations on dependently typed linear sequent calculus (based on Curien & Herbelin's $\mu\widetilde{\mu}$). The current version of his system resembles Andreoli's focalised linear logic (yet to be published).

Pierre-Marie Pédrot has been working on a delimited CPS translation of the Calculus of Inductive Constructions, seen through the prism of polarised linear logic. Restricting dependencies to positives naturally fits into the scenery of delimited control, while extending negatives to infinitary objects permits to recover some properties of the involutivity of linear double-negation.

6.1.4. Proving with side-effects

Axiom of dependent choice. Hugo Herbelin showed that classical arithmetic in finite types extended with strong elimination of existential quantification proves the axiom of dependent choice. To get classical logic and choice together without being inconsistent is made possible first by constraining strong elimination of existential quantification to proofs that are essentially intuitionistic and secondly by turning countable universal quantification into an infinite conjunction of classical proofs evaluated along a call-by-need evaluation strategy so as to extract from them intuitionistic contents that complies to the intuitionistic constraint put on strong elimination of existential quantification. This work has been presented at LICS 2012 [22].

Memory assignment, forcing and delimited control. Hugo Herbelin investigated how to extend his work on intuitionistically proving Markov's principle [54] and the work of Danko Ilik on intuitionistically proving the double negation shift (i.e. $\forall x\, \neg\neg A \to \neg\neg \forall x\, A$) [15] to other kind of effects. In particular, memory assignment is related to Cohen's forcing as emphasised by Krivine [58] and by the observation that Cohen's translation of formula P into $\forall y \leq x \exists z \leq y\, P(z)$ is similar to a state-passing-style transformation of type P into $S \to S \times P$.

Hugo Herbelin then designed a logical formalism with memory assignment that allows to *prove* in direct-style any statement provable using the forcing method, the same way as logic extended with control operators allows to support direct-style classical reasoning. Thanks to the use of delimiters over "small" formulas similar to the notation of Σ^0_1-formulas in arithmetic, the whole framework remains intuitionistic, in the sense that it satisfies the disjunction and existence property.

Two typical applications of proving with side-effects are global-memory proofs of the axiom of countable choice and an enumeration-free proof of Gödel's completeness theorem.

The main ideas of this research program have been communicated during the Logic and Interaction weeks in Marseille in February 2012.

In the continuation of his work with Silvia Ghilezan [4] on showing that Saurin's variant $\Lambda\mu$ [8] of Parigot's $\lambda\mu$-calculus [65] for classical logic was a canonical call-by-name version of Danvy-Filinski's call-by-value calculus of delimited control, Hugo Herbelin studied with Alexis Saurin and Silvia Ghilezan another variant of call-by-name calculus of delimited control. This is leading to a general paper on call-by-name and call-by-value delimited control.

Classical logic, stack calculus and stream calculus. Alexis Saurin studied the connection between the stack calculus recently proposed by Ehrhard et al and $\lambda\mu$-calculus and how the former can be precisely compared to the target of the CPS of the latter. He also investigated separation issues related to the stack calculus. During a visit to UPenn in the spring, Alexis Saurin and Marco Gaboardi investigated type systems for a stream calculus which contains $\Lambda\mu$.

Moreover, Alexis Saurin's paper *Böhm theorem and Böhm trees for the $\Lambda\mu$-calculus* [16] was published in TCS early 2012.

6.1.5. PTS and delimited control

From the study of one-pass CPS on the one side and of previous presentations of pure type systems with control operators on the other side, Pierre Boutillier and Hugo Herbelin have investigated how splitting terms into categories opens a new way to merge dependent types and monads. A preliminary set of rules has been presented during the third week of Logic and Interaction in Marseille.

It was refined since then but has not reached yet the maturity required to be accepted for publication in an international conference.

6.1.6. Interactive realisability

Thanks to the Curry-Howard correspondence for classical logic, it is possible to extract programs from classical proofs. These programs use control operators as a way to implement backtracking and processes of intelligent learning by trial and error. Unfortunately, such programs tend to be poorly efficient. The reason is that, in a sense, they are designed in order to keep their correctness and termination proofs simple. Each small modification of these programs seems, at best, to require major and difficult adaptations of their correctness proofs. This is due to a lack of understanding and control of the backtracking mechanism that interprets classical proofs. In order to write down more efficient programs, it is necessary to describe exactly: a) what the programs learn, b) how the knowledge of programs varies during the execution.

A first step towards this goal is the theory of Interactive Realisability, a semantics for intuitionistic arithmetic with excluded middle over semi-decidable predicates. It is based on a notion of state, which describes the knowledge of programs coming from a classical proof, and explains how the knowledge evolves during computation.

Federico Aschieri has extended the theory of interactive realisability to a full classical system, namely first-order Peano arithmetic with Skolem axioms. This is a very expressive system, with non-trivial axioms of choice and comprehension. The resulting programs are interpreted as stratified-learning algorithms, which build in a very organised way the approximations of the Skolem functions used in the proofs. The work has appeared in the proceedings of the conference Computer Science Logic 2012. A careful implementation of this extended theory –yet to be developed – will lead to a dramatic efficiency improvement over the already existing computational interpretations.

Federico Aschieri has also showed how to use interactive realisability to provide purely proof-theoretic results. He proved with a new method the conservativity of Peano arithmetic with Skolem axioms over Peano arithmetic alone for arithmetical formulas. In particular, the method can be seen as a constructivisation and substantial refinement of Avigad's forcing. The work has appeared in the proceedings of the workshop Classical Logic and Computation 2012.

6.1.7. Reverse mathematics

Hugo Herbelin explored with Gyesik Lee and Keiko Nakata the constructive content of the big five subsystems of second-order arithmetic considered in the context of (classical) reverse mathematics. They obtained a uniform characterisation of these systems in terms of variants of the comprehension axiom called separation, co-separation and interpolation.

This is the first step in a larger project attempting first to connect to predicative type theory the subsystems of System F underlying the proof-as-program structure of the big five subsystems of second-order arithmetic, and secondly to reformulate these subsystems in terms of pure systems of inductive definitions.

Jaime Gaspar has several projects running simultaneously. For example, in one of his projects he created a small unoptimised automated theorem prover, and he hopes to optimise it and use it to obtain a certain completely formalised proof to which he can apply a proof interpretation in order to extract computational content. As another example, in another project he his trying to show that several classical proof interpretations are instances of a unified proof interpretation, in a parallel way to what is known for intuitionistic proof interpretations.

6.2. Type theory and the foundations of Coq

Participants: Pierre Boutillier, Pierre-Louis Curien, Hugo Herbelin, Pierre-Marie Pédrot, Yann Régis-Gianas, Alexis Saurin, Matthieu Sozeau.

6.2.1. Calculus of inductive constructions and typed equality

The work of Hugo Herbelin and Vincent Siles on the equivalence of Pure Type Systems with typed or untyped equality has been published [17].

6.2.2. Substitutions and isomorphisms

Pierre-Louis Curien completed his joint work with Martin Hofmann (Univ. of Munich) and Richard Garner (MacQuarie University, Sydney) on comparing two categorical interpretations of (extensional) type theories. More precisely, we wanted to compare two ways of giving a categorical interpretation of Martin-Löf type theory, both overcoming the following mismatch: syntax has exact substitutions, while their categorical interpretation, in terms of pullbacks or fibrations, “implements” substitutions only up to isomorphism. One can then either change the model (strictification) [55], or modify the syntax (by introducing explicit substitutions and more importantly explicit coercions between types that are now only isomorphic) [2]. In the latter case, one has to prove a coherence theorem to show that the interpretation is in the end independent from these coercion decorations. Such a proof was given in [2], using rewriting methods. These approaches turn out to be related through a general machinery that relates three kinds of categories, with strict or non strict objects and morphisms. As a bonus we get a new, more conceptual proof of coherence. These results are now being written up for a special issue in honour of Glynn Winskel. In further work, we wish to address intensional, and homotopy type-theoretic versions of these coherence problems.

6.2.3. Homotopy type theory

The univalence axiom proposed by Voevodsky states that for any two types to be equal exactly means being of same cardinality. This new axiom for type theory turns to have very interesting consequences for the practical foundations of formal mathematical reasoning: it smoothly implies other axioms such as functional extensionality or propositional existential but before all it says that any property proved about some mathematical structure immediately applies to any other other type (“sets” informally) which it is isomorphic to.

This axiom however contradicts the current logical foundations of Coq (in the presence of Streicher’s axiom K). Investigations have then been started to understand how to weaken the Calculus of Inductive Constructions implemented in Coq so as to make it compatible with univalence. In a first step, this resulted in the design of a new rule for singleton elimination that has been implemented by Hugo Herbelin as an optional feature of Coq (singleton elimination is the ability to build objects in datatypes from canonically-proved propositional properties such as equality).

6.2.4. Models of type theory

The existing models of homotopy type theory are based on simplicial sets or on their extensions as Kan complexes. Hugo Herbelin developed a concrete type-theoretic formalisation of semi-simplicial sets following ideas from Steve Awodey, Peter LeFanu Lumsdaine and other researchers both at Carnegie-Mellon University and at the Institute of Advanced Study. The technique he used seems to straightforwardly generalise to provide type-theoretic constructions for arbitrary presheaves on inductively generated categories.

6.2.5. Forcing in type theory

Together with Nicolas Tabareau and Guilhem Jaber (Inria Ascola team, École des mines de Nantes), Matthieu Sozeau investigated an internalisation of the presheaf model of the Calculus of Inductive Constructions (CIC). They published their work at LICS'12 [23]. This work corresponds to adapting the idea of Forcing due to Cohen in Type Theory. An internal model construction allows to enrich the logical type theory with new modalities and define their semantics by translation to CIC. The usual Cohen forcing can be realised using this framework to show the independence of the continuum hypothesis in CIC, but more practical applications are possible as well. Notably, the step-indexed technique for building models of imperative languages with rich type structure can be phrased as a forcing/presheaf construction. Sozeau, Tabareau and Jaber developed a plugin that can handle this example [32] which relies on a modified Coq version implementing proof-irrelevance and eta-rules for records.

6.2.6. Proof irrelevance, eta-rules

Matthieu Sozeau continued his work on proof-irrelevance by implementing a variant of Werner's proof-irrelevant CIC in Coq [72]. An article describing this work is in preparation. The new system also handles the extensional eta-rules for records, extending the technique implemented by Hugo Herbelin to handle eta-expansion of functions in Coq.

6.2.7. Unification

The unification algorithm of Coq now essentially dwells in the λ-calculus part of the language. Pierre Boutillier started a refactoring of the code in order to deal with algebraic datatypes. Hugo Herbelin and Pierre Boutillier investigated how to reformulate unification on top of an abstract machine (i.e. on top of sequent calculus). Hugo Herbelin added various heuristics to the unification algorithm of Coq, making them both more powerful and customisable.

Matthieu Sozeau is continuing work in collaboration with Beta Ziliani (PhD student of Derek Dreyer at MPI Saarbrücken, two one week visits in 2012), and Aleksandar Nanevski (Researcher at IMDEA Madrid) on giving a clear formalisation for the unification algorithm of Coq. This will help understand better the working of advanced features like Canonical Structures and Type Classes that are heavily used in big developments, as the spectacular recently completed formalisation of the proof of Feit-Thompson's Odd theorem by the Mathematical Components team.

Matthieu Sozeau adapted the existing unification algorithm to be universe-aware, resulting in more predictability and earlier error-reporting in both the type inference and tactic unification algorithms of Coq.

6.3. Homotopy of rewriting systems

Participants: Pierre-Louis Curien, Yves Guiraud, Philippe Malbos.

6.3.1. Coherence in monoidal structures

Yves Guiraud and Philippe Malbos have applied the Squier's homotopical theorem [70], which they had generalised to higher-dimensional rewriting systems [52], to several types of categories with monoidal structures. This work develops a formal setting to produce constructive proofs of coherence conditions, applied to the cases of monoidal categories, symmetric monoidal categories and braided categories. These results have been published in Mathematical Structures in Computer Science [12].

6.3.2. Computation of resolutions of monoids

Yves Guiraud and Philippe Malbos have extended Squier's homotopical theory to the higher dimensions of presentations of monoids to get an algorithm transforming a convergent word rewriting system into a polygraphic resolution of the presented monoid, in the setting of Métayer [63]. From this polygraphic resolution, this work gives an explicit procedure to recover several of the known Abelian resolutions of the monoid, generalising and relating algebraic invariants of monoids. Moreover, a polygraphic resolution corresponds to the normalisation strategies of rewriting systems and they contain all the proofs of equality between elements, together with the proofs of equality of those proofs of equality, and so on. This work has been published in Advances in Mathematics [13]. By nature, polygraphic resolutions bear many similarities with the higher-dimensional groupoids that appear in homotopical type theory when one considers the towers of identity types: this connection will be investigated by Pierre-Louis Curien, Yves Guiraud, Hugo Herbelin and Matthieu Sozeau.

6.3.3. Coherent presentations of Artin groups

With Stéphane Gaussent (Institut Camille Jordan, Université de Saint-Étienne), Yves Guiraud and Philippe Malbos are currently finishing an article on the rewriting properties and coherence issues in Artin groups, a class of groups that is fundamental in algebra and geometry. This work uses the formal setting of coherent presentations (a truncation of polygraphic resolutions at the level above relations) to formulate, in a common language, several known results in combinatorial group theory: one by Tits about the fundamental group of a graph associated to an Artin group [71], and one by Deligne about the actions of Artin groups on categories [44], both proved by geometrical and non-constructive methods. In this work, an improvement of Knuth-Bendix's completion procedure is introduced, called the homotopical completion-reduction procedure, and it is used to give a constructive proof of both those theorems. In fact, the method even improves the formerly known results: for example, it generalises Deligne's result to cases where his geometrical proof cannot be applied. A preliminary version of this work is available online [31]. The next objective of this collaboration is to extend the formal setting and methods to compute polygraphic resolutions of Artin groups, with a view towards two open problems of combinatorial group theory with respect to Artin groups: the decidability of the word problem and the verification that a precise topological space is a classifying space.

6.3.4. Higher-dimensional linear rewriting

With Samuel Mimram (CEA Saclay) and Pierre-Louis Curien, Yves Guiraud and Philippe Malbos investigate the links between set-theoretic rewriting theory and the computational methods known in symbolic algebra, such as Gröbner bases [36]. In particular, this work is interested in extending the setting of higher-dimensional rewriting to include "linear rewriting" and, as a consequence, to be able to apply its methods in symbolic computation. One particular direction is to understand Anick's resolution [33], and to improve it with the completion-reduction methodology, in order to get better algorithms to compute homological invariants and to prove important properties such as Koszulness. This research direction has been presented to the first call for projects of the IDEX Sorbonne-Paris-Cité, together with Eric Hoffbeck and Muriel Livernet (LAGA, Université Paris 13) and François Métayer (PPS, Université Paris 7).

6.4. Coq as a functional programming language

Participants: Nicolas Ayache, Pierre Boutillier, François Bobot, Guillaume Claret, Lourdes del Carmen Gonzalez Huesca, Tim Griffin, Hugo Herbelin, Pierre Letouzey, Matthias Puech, Yann Régis-Gianas, Matthieu Sozeau.

6.4.1. Type classes and libraries

Pierre Castéran from Inria Bordeaux and Matthieu Sozeau published a tutorial on the use of type classes [30] that was used as the basis of an invited lecture by M. Sozeau at the JFLA conference in February 2012. It will be published as part of the new version of the Coq'Art book.

6.4.2. Dependent pattern-matching

Pierre Boutillier experimented about how to integrate gently in Coq the compilation process he came up with last year to simulate Agda-style dependent pattern-matching. As a consequence, pattern grammar in Gallina has changed, much more notations can be used and users can write patterns instead of simple abstractions in the pattern-matching return clause.

Matthieu Sozeau continued maintenance and polishing of the Equations plugin that allows dependent pattern-matching on inductive families. A first official release is planned for the beginning of 2013.

6.4.3. Modularised arithmetical libraries

The modularised arithmetical libraries elaborated by Pierre Letouzey during the previous year(s) have been released as part of Coq 8.4. They provided greater uniformity of available functions and lemmas across the various Coq numerical datatypes. These libraries seem to work quite well, the only remaining issue is the documentation: due to this complex modular organisation, it is currently tedious for the user to browse the available functions and results. We expect to tackle this issue next year, by providing various documentation views, either the external summary of all available elements at once, or the internal layout of these elements.

6.4.4. Library of finite sets

Pierre Letouzey has integrated an additional Coq implementation in the MSets library of finite sets. This additional implementation is an improved version of the Red-Black-Tree library contributed by Andrew Appel. Using these RBT instead of the previously available AVL could be more efficient, at least in Coq, since they trigger no computations of integer numbers coding the tree depth.

6.4.5. Library on XPath processing

As part of the ANR Typex (http://typex.lri.fr), Matthieu Sozeau is developing a library for the certification of efficient XPath/XQuery engines in collaboration with Kim Nguyen (LRI) and Alan Schmitt (Inria Grenoble).

6.4.6. Mathematics of routing

Tim Griffin's primary focus during his visit to πr^2 was the development of a “metalanguage" for algebraic structures using Coq. Since he was something of a beginner with Coq, this involved learning the basics as well as more advanced work on representing algebraic structures. He made very good progress on this while in Paris and is now continuing this work in Cambridge.

6.4.7. Incrementality in proof languages

Matthias Puech and Yann Régis-Gianas worked on incremental type checking. This preliminary work has been presented during a contributed talk at TLDI 2012 [25]. It sets the ground for an incremental proof development and checking system, by means of a representation language for repositories of proofs and proof changes.

The traditional interaction with a proof-checker is a batch process. Coq (among others) refines this model by providing a read-eval print loop with a global undo system, implemented in an ad-hoc way. A more general approach to incrementality is being developed by means of a finer-grained analysis of dependencies. The approach developed is adaptable to any typed formal language: the language is specified in a meta-language close to the Edinburgh Logical Framework, in which subsequent versions of a development can be type-checked incrementally. Applications of this framework are: proof language for proof assistants, integrated development environments for proof or programming languages, typed version control systems.

6.4.8. Proofs of programs in Coq

As part of the CerCo European project, in collaboration with Roberto Amadio (PPS, Paris 7), François Bobot, Nicolas Ayache and Yann Régis-Gianas maintained a prototype compiler for a large subset of the C language whose specificity is to annotate input source programs with information about the worst-case execution cost of their machine code. Yann Régis-Gianas started a mechanised version of the proof technique used to prove the correctness of such an annotating compiler.

Yann Régis-Gianas maintained another compiler for Core ML that uses a generalisation of CerCo technique in order to obtain certified worst case execution time bounds on functional programs. This compiler produces proof obligations in Coq. The corresponding paper is published in January 2012 in the proceedings of FOPARA 2011 [19].

Nicolas Ayache developed a Frama-C plugin distributed in the CerCo software suite whose role is to synthesize cost annotations out of C programs. François Bobot developed a new version of this plugin. In particular, this new version handles C programs that manipulate pointers.

In collaboration with Roberto Amadio, Yann Régis-Gianas extended the cost annotating compilation chain of the FOPARA paper to handle the cost of memory management. A journal paper is about to be published in HOSC.

6.4.9. Lightweight proof-by-reflection

In the context of the ANR project Paral-ITP, Lourdes del Carmen Gonzalez Huesca, Guillaume Claret and Yann Régis-Gianas developed a new technique for proof-by-reflection based on a notion of *a posteriori* simulation of effectful computations in Coq.

POLSYS Project-Team

6. New Results

6.1. The complexity of solving quadratic boolean systems is better than exhaustive search

A fundamental problem in computer science is to find all the common zeroes of m quadratic polynomials in n unknowns over $\mathbb{F}_2$. The cryptanalysis of several modern ciphers reduces to this problem. Up to now, the best complexity bound was reached by an exhaustive search in $4\log_2 n\, 2^n$ operations. In [4], we give an algorithm that reduces the problem to a combination of exhaustive search and sparse linear algebra. This algorithm has several variants depending on the method used for the linear algebra step. Under precise algebraic assumptions on the input system, we show in [4] that the deterministic variant of our algorithm has complexity bounded by $O(2^{0.841n})$ when $m=n$, while a probabilistic variant of the Las Vegas type has expected complexity $O(2^{0.792n})$. Experiments on random systems show that the algebraic assumptions are satisfied with probability very close to 1. We also give a rough estimate for the actual threshold between our method and exhaustive search, which is as low as 200, and thus very relevant for cryptographic applications.

6.2. Improving the Complexity of Index Calculus Algorithms in Elliptic Curves over Binary Fields

In [25], we study the index calculus method that was first introduced by Semaev for solving the ECDLP and later developed by Gaudry and Diem. In particular, we focus on the step which consists in decomposing points of the curve with respect to an appropriately chosen factor basis. This part can be nicely reformulated as a purely algebraic problem consisting in finding solutions to a multivariate polynomial. Our main contribution is the identification of particular structures inherent to such polynomial systems and a dedicated method for tackling this problem. We solve it by means of Gröbner basis techniques and analyze its complexity using the multi-homogeneous structure of the equations. We emphasize that the complexity obtained in the paper is very conservative in comparison to experimental results. We hope the new ideas provided here may lead to efficient index calculus based methods for solving ECDLP in theory and practice.

6.3. On the relation between the MXL family of algorithms and Gröbner basis algorithms

The computation of Gröbner bases remains one of the most powerful methods for tackling the Polynomial System Solving (PoSSo) problem. The most efficient known algorithms reduce the Gröbner basis computation to Gaussian eliminations on several matrices. However, several degrees of freedom are available to generate these matrices. It is well known that the particular strategies used can drastically affect the efficiency of the computations. In this work, we investigate a recently-proposed strategy, the so-called “Mutant strategy”, on which a new family of algorithms is based (MXL, MXL2 and MXL3). By studying and describing the algorithms based on Gröbner basis concepts, we demonstrate in [3] that the Mutant strategy can be understood to be equivalent to the classical Normal Selection Strategy currently used in Gröbner basis algorithms. Furthermore, we show that the “partial enlargement” technique can be understood as a strategy for restricting the number of S-polynomials considered in an iteration of the F4 Gröbner basis algorithm, while the new termination criterion used in MXL3 does not lead to termination at a lower degree than the classical Gebauer–Möller installation of Buchberger’s criteria. We claim that our results map all novel concepts from the MXL family of algorithms to their well-known Gröbner basis equivalents. Using previous results that had shown the relation between the original XL algorithm and F4, we conclude that the MXL family of algorithms can be fundamentally reduced to redundant variants of F4.

6.4. On the Complexity of the BKW Algorithm on LWE

In [35], we present a study of the complexity of the Blum-Kalai-Wasserman (BKW) algorithm when applied to the Learning with Errors (LWE) problem, by providing refined estimates for the data and computational effort requirements for solving concrete instances of the LWE problem. We apply this refined analysis to suggested parameters for various LWE-based cryptographic schemes from the literature and, as a result, provide new upper bounds for the concrete hardness of these LWE-based schemes.

6.5. On the Complexity of the Arora-Ge algorithm against LWE

Arora & Ge recently showed that solving LWE can be reduced to solve a high-degree non-linear system of equations. They used a linearization to solve the systems. We investigate in [34] the possibility of using Gröbner bases to improve Arora & Ge approach.

6.6. On enumeration of polynomial equivalence classes and their application to MPKC

The Isomorphism of Polynomials (IP) is one of the most fundamental problems in multivariate public key cryptography (MPKC). In [8], we introduce a new framework to study the counting problem associated to IP. Namely, we present tools of finite geometry allowing to investigate the counting problem associated to IP. Precisely, we focus on enumerating or estimating the number of isomorphism equivalence classes of homogeneous quadratic polynomial systems. These problems are equivalent to finding the scale of the key space of a multivariate cryptosystem and the total number of different multivariate cryptographic schemes respectively, which might impact the security and the potential capability of MPKC. We also consider their applications in the analysis of a specific multivariate public key cryptosystem. Our results not only answer how many cryptographic schemes can be derived from monomials and how big the key space is for a fixed scheme, but also show that quite many HFE cryptosystems are equivalent to a Matsumoto-Imai scheme.

6.7. Cryptanalysis of HFE, Multi-HFE and Variants for Odd and Even Characteristic

In [5], we investigate the security of HFE and Multi-HFE schemes as well as their minus and embedding variants. Multi-HFE is a generalization of the well-known HFE schemes. The idea is to use a multivariate quadratic system - instead of a univariate polynomial in HFE - over an extension field as a private key. According to the authors, this should make the classical direct algebraic (message-recovery) attack proposed by Faugère and Joux on HFE no longer efficient against Multi-HFE. We consider here the hardness of the key-recovery in Multi-HFE and its variants, but also in HFE (both for odd and even characteristic). We first improve and generalize the basic key recovery proposed by Kipnis and Shamir on HFE. To do so, we express this attack as matrix/vector operations. In one hand, this permits to improve the basic Kipnis-Shamir (KS) attack on HFE. On the other hand, this allows to generalize the attack on Multi-HFE. Due to its structure, we prove that a Multi-HFE scheme has much more equivalent keys than a basic HFE. This induces a structural weakness which can be exploited to adapt the KS attack against classical modifiers of multivariate schemes such as minus and embedding. Along the way, we discovered that the KS attack as initially described cannot be applied against HFE in characteristic 2. We have then strongly revised KS in characteristic 2 to make it work. In all cases, the cost of our attacks is related to the complexity of solving MinRank. Thanks to recent complexity results on this problem, we prove that our attack is polynomial in the degree of the extension field for all possible practical settings used in HFE and Multi-HFE. This makes then Multi-HFE less secure than basic HFE for equally-sized keys. As a proof of concept, we have been able to practically break the most conservative proposed parameters of multi-HFE in few days (256 bits security broken in 9 days).

6.8. Solving Polynomial Systems over Finite Fields: Improved Analysis of the Hybrid Approach

The Polynomial System Solving (PoSSo) problem is a fundamental NP-Hard problem in computer algebra. Among others, PoSSo have applications in area such as coding theory and cryptology. Typically, the security of multivariate public-key schemes (MPKC) such as the UOV cryptosystem of Kipnis, Shamir and Patarin is directly related to the hardness of PoSSo over finite fields. The goal of [22] is to further understand the influence of finite fields on the hardness of PoSSo. To this end, we consider the so-called *hybrid approach*. This is a polynomial system solving method dedicated to finite fields proposed by Bettale, Faugère and Perret (Journal of Mathematical Cryptography, 2009). The idea is to combine exhaustive search with Gröbner bases. The efficiency of the hybrid approach is related to the choice of a trade-off between the two methods. We propose here an improved complexity analysis dedicated to quadratic systems. Whilst the principle of the hybrid approach is simple, its careful analysis leads to rather surprising and somehow unexpected results. We prove that the optimal trade-off (i.e. number of variables to be fixed) allowing to minimize the complexity is achieved by fixing a number of variables proportional to the number of variables of the system considered, denoted n. Under some natural algebraic assumption, we show that the asymptotic complexity of the hybrid approach is $2^{(3.31-3.62\,\log_2(q)^{-1})\,n}$, where q is the size of the field (under the condition in particular that $\log(q) \ll n$). This is to date, the best complexity for solving PoSSo over finite fields (when $q > 2$). We have been able to quantify the gain provided by the hybrid approach compared to a direct Gröbner basis method. For quadratic systems, we show (assuming a natural algebraic assumption) that this gain is exponential in the number of variables. Asymptotically, the gain is $2^{1.49\,n}$ when both n and q grow to infinity and $\log(q)$.

6.9. Efficient Arithmetic in Successive Algebraic Extension Fields Using Symmetries

In [15] we present new results for efficient arithmetic operations in a number field $\mathbb{K}$ represented by successive extensions. These results are based on multi-modular and evaluation–interpolation techniques. We show how to use intrinsic symmetries in order to increase the efficiency of these techniques. Applications to splitting fields of univariate polynomials are presented.

6.10. Algebraic Crypanalysis with Side Channels Information

In [6] and [24] (see also the PhD thesis of C. Goyet [1]), we present new cryptanalyses of symmetric and asymmetric cryptosystems (e.g. AES and ECDSA). These analyses share the same fundamental hypotheses that some information are provided to the attacker by some oracle. In a practical point of view, such an oracle can be represented as a partial side channel attack realized in a first step (e.g. SPA, Fault attacks). The second step of the attack uses algorithms from computer algebra (e.g. Gröbner basis computation, LLL) in order to retrieve the secret key.

6.11. Worst case complexity of the Continued Fraction (CF) algorithm.

In [16] we consider the problem of isolating the real roots of a square-free polynomial with integer coefficients using the classic variant of the continued fraction algorithm (CF), introduced by Akritas. We compute a lower bound on the positive real roots of univariate polynomials using exponential search. This allows us to derive a worst case bound of $\widetilde{O}(d^4\tau^2)$ for isolating the real roots of a polynomial with integer coefficients using the *classic variant of CF*, where d is the degree of the polynomial and τ the maximum bitsize of its coefficients. This improves the previous bound of Sharma by a factor of d^3 and matches the bound derived by Mehlhorn and Ray for another variant of CF which is combined with subdivision; it also matches the worst case bound of the classical subdivision-based solvers STURM, DESCARTES, and BERNSTEIN.

6.12. Local Generic Position for Root Isolation of Zero-dimensional Triangular Polynomial Systems.

In [30] we present an algorithm based on local generic position (LGP) to isolate the complex or real roots and their multiplicities of a zero-dimensional triangular polynomial system. The Boolean complexity of the algorithm for computing the real roots is single exponential: $\widetilde{O}_B(N^{n^2})$, where $N = \max\{d, \tau\}$, d and τ, is the degree and the maximum coefficient bitsize of the polynomials, respectively, and n is the number of variables.

6.13. Univariate Real Root Isolation in Multiple Extension Fields

In [31] we present algorithmic, complexity and implementation results for the problem of isolating the real roots of a univariate polynomial in $K[x] \in L[y]$, where $L = \mathbb{Q}(\alpha\_1, \cdots, \alpha\_\ell)$ is an algebraic extension of the rational numbers. Our bounds are single exponential in ℓ and match the ones presented for the case $\ell = 1$. We consider two approaches. The first, indirect approach, using multivariate resultants, computes a univariate polynomial with integer coefficients, among the real roots of which are the real roots of B_α. The Boolean complexity of this approach is $\widetilde{O}_B(N^{4\ell+4})$, where N is the maximum of the degrees and the coefficient bitsize of the involved polynomials. The second, direct approach, tries to solve the polynomial directly, without reducing the problem to a univariate one. We present an algorithm that generalizes Sturm algorithm from the univariate case, and modified versions of well known solvers that are either numerical or based on Descartes' rule of sign. We achieve a Boolean complexity of $\widetilde{O}_B(\min\{N^{4\ell+7}, N^{2\ell^2+6}\})$ and $\widetilde{O}_B(N^{2\ell+4})$, respectively. We implemented the algorithms in C as part of the core library of Mathematica and we illustrate their efficiency over various data sets.

6.14. Mixed volume and distance geometry techniques for counting Euclidean embeddings of rigid graphs.

A graph G is called generically minimally rigid in $\mathbb{R}^d$ if, for any choice of sufficiently generic edge lengths, it can be embedded in $\mathbb{R}^d$ in a finite number of distinct ways, modulo rigid transformations. In [37] we deal with the problem of determining tight bounds on the number of such embeddings, as a function of the number of vertices. The study of rigid graphs is motivated by numerous applications, mostly in robotics, bioinformatics, and architecture. We capture embeddability by polynomial systems with suitable structure, so that their mixed volume, which bounds the number of common roots, yields interesting upper bounds on the number of embeddings. We explore different polynomial formulations so as to reduce the corresponding mixed volume, namely by introducing new variables that remove certain spurious roots, and by applying the theory of distance geometry. We focus on $\mathbb{R}^2$ and $\mathbb{R}^3$, where Laman graphs and 1-skeleta of convex simplicial polyhedra, respectively, admit inductive Henneberg constructions. Our implementation yields upper bounds for $n \leq 10$ in $\mathbb{R}^2$ and $\mathbb{R}^3$, which reduce the existing gaps and lead to tight bounds for $n \leq 7$ in both $\mathbb{R}^2$ and $\mathbb{R}^3$; in particular, we describe the recent settlement of the case of Laman graphs with 7 vertices. We also establish the first lower bound in $\mathbb{R}^3$ of about 2.52^n, where n denotes the number of vertices.

6.15. Variant Quantifier Elimination

In [10], we describe an algorithm (VQE) for a *variant* of the real quantifier elimination problem (QE). The variant problem requires the input to satisfy a certain *extra condition*, and allows the output to be *almost* equivalent to the input. The motivation/rationale for studying such a variant QE problem is that many quantified formulas arising in applications do satisfy the extra conditions. Furthermore, in most applications, it is sufficient that the output formula is almost equivalent to the input formula. The main idea underlying the algorithm is to substitute the repeated projection step of CAD by a single projection without carrying out a parametric existential decision over the reals. We find that the algorithm can tackle important and challenging problems, such as numerical stability analysis of the widely-used MacCormack's scheme. The problem has been practically out of reach for standard QE algorithms in spite of many attempts to tackle it. However the current implementation of VQE can solve it in about 12 hours.

6.16. Global optimization

Let $f_1, \cdots, f_p$ be in $\mathbb{Q}[\mathbf{X}]$, where $\mathbf{X} = (X_1, \cdots, X_n)^t$, that generate a radical ideal and let V be their complex zero-set. Assume that V is smooth and equidimensional. Given $f \in \mathbb{Q}[X]$ bounded below, consider the optimization problem of computing $f^{\star} = \inf_{x \in V \cap \mathbb{R}^n} f(x)$. For $\mathbf{A} \in GL_n(\mathbb{C})$, we denote by $f^{\mathbf{A}}$ the polynomial $f(\mathbf{AX})$ and by $V^{\mathbf{A}}$ the complex zero-set of $f_1^{\mathbf{A}}, ..., f_p^{\mathbf{A}}$. In [9], we construct families of polynomials $\mathsf{M}_0^{\mathbf{A}}, ..., \mathsf{M}_d^{\mathbf{A}}$ in $\mathbb{Q}[\mathbf{X}]$: each $\mathsf{M}_i^{\mathbf{A}}$ is related to the section of a linear subspace with the critical locus of a linear projection. We prove that there exists a non-empty Zariski-open set $O \subset GL_n(\mathbb{C})$ such that for all $\mathbf{A} \in O \cap GL_n(\mathbb{Q})$, $f(x)$ is non-negative for all $x \in V \cap \mathbb{R}^n$ if, and only if, $f^{\mathbf{A}}$ can be expressed as a sum of squares of polynomials on the truncated variety generated by the ideal $\langle \mathsf{M}_i^{\mathbf{A}} \rangle$, for $0 \leq i \leq d$. Hence, we can obtain algebraic certificates for lower bounds on $f^{\star}$ using semidefinite programs. Some numerical experiments are given. We also discuss how to decrease the number of polynomials in $\mathsf{M}_i^{\mathbf{A}}$.

6.17. Gröbner bases and critical points

We consider the problem of computing critical points of the restriction of a polynomial map to an algebraic variety. This is of first importance since the global minimum of such a map is reached at a critical point. Thus, these points appear naturally in non-convex polynomial optimization which occurs in a wide range of scientific applications (control theory, chemistry, economics,etc.). Critical points also play a central role in recent algorithms of effective real algebraic geometry. Experimentally, it has been observed that Gröbner basis algorithms are efficient to compute such points. Therefore, recent software based on the so-called Critical Point Method are built on Gröbner bases engines. Let $f_1, ..., f_p$ be polynomials in $\mathbb{Q}[x_1, ..., x_n]$ of degree D, $V \subset \mathbb{C}^n$ be their complex variety and π_1 be the projection map $(x_1, ..., x_n) \mapsto x_1$. The critical points of the restriction of π_1 to V are defined by the vanishing of $f_1, ..., f_p$ and some maximal minors of the Jacobian matrix Indus associated to $f_1, ..., f_p$. Such a system is algebraically structured: the ideal it generates is the sum of a determinantal ideal and the ideal generated by $f_1, ..., f_p$. In [26], we provide the first complexity estimates on the computation of Gröbner bases of such systems defining critical points. We prove that under genericity assumptions on $f_1, ..., f_p$, the complexity is polynomial in the generic number of critical points, i.e.$D^p(D-1)^{n-p} \begin{pmatrix} n-1 \\ p-1 \end{pmatrix}$ More particularly, in the quadratic case $D = 2$, the complexity of such a Gröbner basis computation is polynomial in the number of variables n and exponential in p. We also give experimental evidence supporting these theoretical results.

POP ART Project-Team

6. New Results

6.1. Dependable Distributed Real-time Embedded Systems

Participants: Gwenaël Delaval, Pascal Fradet, Alain Girault [contact person], Emil Dumitrescu.

6.1.1. Tradeoff exploration between reliability, power consumption, and execution time

For autonomous critical real-time embedded systems (*e.g.*, satellite), guaranteeing a very high level of reliability is as important as keeping the power consumption as low as possible. We have designed an off-line ready list scheduling heuristics which, from a given software application graph and a given multiprocessor architecture (homogeneous and fully connected), produces a static multiprocessor schedule that optimizes three criteria: its *length* (crucial for real-time systems), its *reliability* (crucial for dependable systems), and its *power consumption* (crucial for autonomous systems). Our tricriteria scheduling heuristics, **TSH**, uses the *active replication* of the operations and the data-dependencies to increase the reliability, and uses *dynamic voltage and frequency scaling* to lower the power consumption [17], [11]. By running TSH on a single problem instance, we are able to provide the Pareto front for this instance in 3D, therefore exposing the user to several tradeoffs between the power consumption, the reliability and the execution time. The new contribution for 2012 has been the formulation of a new multi-criteria cost function for our ready list scheduling heuristics, such that we are able to prove rigorously that the static schedules we generate meet both the reliability constraint and the power consumption constraint.

Thanks to extensive simulation results, we have shown how TSH behaves in practice. Firstly, we have compared TSH versus an optimal Mixed Linear Integer Program on small instances; the experimental results show that TSH behaves very well compared to the the ILP. Secondly, we have compared TSH versus the ECS heuristic (Energy-Conscious Scheduling [68]); the experimental results show that TSH performs systematically better than ECS.

This is a joint work with Ismail Assayad (U. Casablanca, Morocco) and Hamoudi Kalla (U. Batna, Algeria), who both visit the team regularly.

6.2. Controller Synthesis for the Safe Design of Embedded Systems

Participants: Gwenaël Delaval [contact person], Gregor Goessler, Sebti Mouelhi.

6.2.1. Synthesis of switching controllers using approximately bisimilar multiscale abstractions

The use of discrete abstractions for continuous dynamics has become standard in hybrid systems design (see *e.g.*, [73] and the references therein). The main advantage of this approach is that it offers the possibility to leverage controller synthesis techniques developed in the areas of supervisory control of discrete-event systems [71]. The first attempts to compute discrete abstractions for hybrid systems were based on traditional systems behavioral relationships such as simulation or bisimulation, initially proposed for discrete systems most notably in the area of formal methods. These notions require inclusion or equivalence of observed behaviors which is often too restrictive when dealing with systems observed over metric spaces. For such systems, a more natural abstraction requirement is to ask for closeness of observed behaviors. This leads to the notions of approximate simulation and bisimulation introduced in [54].

These notions enabled the computation of approximately equivalent discrete abstractions for several classes of dynamical systems, including nonlinear control systems with or without disturbances, and switched systems. These approaches are based on sampling of time and space where the sampling parameters must satisfy some relation in order to obtain abstractions of a prescribed precision. In particular, the smaller the time sampling parameter, the finer the lattice used for approximating the state-space; this may result in abstractions with a very large number of states when the sampling period is small. However, there are a number of applications where sampling has to be fast; though this is generally necessary only on a small part of the state-space.

In [45] we have proposed a technique for the synthesis of safety controllers for switched systems using multi-scale abstractions that allow us to deal with fast switching while keeping the number of states in the abstraction at a reasonable level. The finest scales of the abstraction are effectively explored only when fast switching is needed, that is when the system approaches the unsafe set.

We have extended the approach of [45] to the synthesis of controllers for time-bounded reachability. Furthermore we have implemented the algorithms for safety and time-bounded reachability in COSYMA, a tool for automatic controller synthesis for incrementally stable switched systems based on multi-scale discrete abstractions. The tool accepts a description of a switched system represented by a set of differential equations and the sampling parameters used to define an approximation of the state-space on which discrete abstractions are computed. The tool generates a controller — if it exists — for the system that enforces a given safety or time-bounded reachability specification.

We are currently exploring, in the SYMBAD project, controller synthesis for switched systems based on a different approach for the construction of multi-scale abstractions. The goal is to further improve the trade-off between cost and precision.

6.2.2. Modular discrete controller synthesis

Discrete controller synthesis (DCS) [71] allows to design programs in a mixed imperative/declarative way. From a program with some freedom degrees left by the programmer (*e.g.*, free controllable variables), and a temporal property to enforce which is not *a priori* verified by the initial program, DCS tools compute off-line automatically a *controller* which will constrain the program (by *e.g.*, giving values to controllable variables) such that, whatever the values of inputs from the environment, the *controlled program* satisfies the temporal property.

Our motivation *w.r.t.* DCS concerns its modular application, improving the scalability of the technique by using contract enforcement and abstraction of components. Moreover, our aim is to integrate DCS into a compilation chain, and thereby improve its usability by programmers, not experts in discrete control. This work has been implemented into the HEPTAGON/BZR language and compiler [50]. This work is done in collaboration with Hervé Marchand (VERTECS team from Rennes) and Eric Rutten (SARDES team from Grenoble).

The implemented tool allows the generation of the synthesized controller under the form of an HEPTAGON node, which can in turn be analyzed and compiled, together with the HEPTAGON source from which it has been generated. This full integration allows this method to aim different target languages (currently C, JAVA or VHDL), and its integrated use in different contexts.

A formal semantics of BZR has been defined, taking into account its underlying nondeterminism related to the presence of controllable variables. A new implementation has been achieved, including an abstraction method based on [47]. We have used BZR for demonstrating the use of Control Theory and Techniques to the administration of computing systems in a closed-loop management [19].

6.3. Automatic Distribution of Synchronous Programs

Participants: Gwenaël Delaval [contact person], Alain Girault, Gregor Goessler, Xavier Nicollin, Gideon Smeding.

6.3.1. Modular distribution

Synchronous programming languages describe functionally centralized systems, where every value, input, output, or function is always directly available for every operation. However, most embedded systems are nowadays composed of several computing resources. The aim of this work is to provide a language-oriented solution to describe *functionally distributed reactive systems*. This research is conducted within the Inria large scale action SYNCHRONICS and is a joint work with Marc Pouzet (ENS, PARKAS team from Rocquencourt) and Xavier Nicollin (Grenoble INP, VERIMAG lab).

We are working on type systems to formalize, in a uniform way, both the clock calculus and the location calculus of a synchronous data-flow programming language (the HEPTAGON language, inspired from LUCID SYNCHRONE [38]). On one hand, the clock calculus infers the clock of each variable in the program and checks the clock consistency: *e.g.*, a time-homogeneous function, like `+`, should be applied to variables with identical clocks. On the other hand, the location calculus infers the spatial distribution of computations and checks the spatial consistency: *e.g.*, a centralized operator, like `+`, should be applied to variables located at the same location. Compared to the PhD of Gwenaël Delaval [48], [49], the goal is to achieve *modular* distribution. By modular, we mean that we want to compile each function of the program into a single function capable of running on any computing location. We make use of our uniform type system to express the computing locations as first-class abstract types, exactly like clocks, which allows us to compile a typed variable (typed by both the clock and the location calculi) into `if ... then ... else ...` structures, whose conditions will be valuations of the clock and location variables.

We currently work on an example of software-defined radio. We have shown on this example how to use a modified clock calculus to describe the localisation of values as clocks, and the architecture as clocks (for the computing resources) and their relations (for communication links).

6.3.2. Distribution of synchronous programs under real-time constraints

With the objective to distribute synchronous data-flow programs (*e.g.*, LUSTRE) over GALS architectures, such that the difference between the original and synchronous systems satisfy given bounds, we have developed a quantitative clock calculus to (1) describe timing properties of the architecture's clock domain, and (2) describe the acceptable difference between the original and distributed programs. The clock calculus is inspired by the network calculus [67], with the difference that clocks are described only with respect to one-another, not with respect to real-time.

As a first result, we have applied our clock calculus to analyze the properties of periodic synchronous data-flow programs executed on a network of processors. Because our clock calculus is relational, it can model and preserve correlated variations of streams. In particular, the common case of a data-flow system that splits a stream for separate treatment, and joins them afterwards, this analysis yields more precise result than comparable methods [24].

We have been able to use the clock calculus as an abstract domain to perform abstract interpretation of synchronous boolean data-flow programs and their distribution on synchronous nodes that communicate asynchronously by sampling shared memory. The analysis discovers the relative clock drift of all clocks of the distributed system as well as bounds on the distance from the original program.

In case the guaranteed maximal distance is too large, we provide methods to synthesize bounds on the relative drift of the architecture's clocks that ensure an acceptable distance. Given the synthesized bounds, we use the known clock drifts and program behavior to synthesize light weight protocols.

6.4. New Programming Languages for Embedded Systems

Participants: Alain Girault [contact person], Pascal Fradet, Vagelis Bebelis, Bertrand Jeannet, Peter Schrammel.

6.4.1. Analysis and scheduling of parametric dataflow models

Recent data-flow programming environments support applications whose behavior is characterized by dynamic variations in resource requirements. The high expressive power of the underlying models (*e.g.*, Kahn Process Networks, the CAL actor language) makes it challenging to ensure predictable behavior. In particular, checking *liveness* (*i.e.*, no part of the system will deadlock) and *boundedness* (*i.e.*, the system can be executed in finite memory) is known to be hard or even undecidable for such models. This situation is troublesome for the design of high-quality embedded systems.

We have introduced the *schedulable parametric data-flow (SPDF)* MoC for dynamic streaming applications [20]. SPDF extends the standard dataflow model by allowing rates to be parametric (*e.g.*, of the form $2xy$, where x and y are parameters, the value of which can change at run-time). SPDF was designed to be statically analyzable while retaining sufficient expressive power. We formulated sufficient and general static criteria for boundedness and liveness. In SPDF, parameters can be changed dynamically even within iterations. The safety of dynamic parameter changes can be checked and their implementation made explicit in the graph. These different analyses are made possible using well-defined static operations on symbolic expressions. The same holds for quasi-static scheduling which is the first step towards code generation for multi-core systems.

We are now focusing on the parallel scheduling of parametric dataflow models. We have proposed a generic and flexible framework to generate parallel ASAP schedules targeted to the new STHORM many-core platform designed by STMicroelectronics. The parametric dataflow graph is associated with generic or user-defined specific constraints aimed at minimizing, timing, buffer sizes, power consumption, or other criteria. The scheduling algorithm executes with minimal overhead and can be adapted to different scheduling policies just by changing some constraints. The safety of both the dataflow graph and constraints can be checked statically and all schedules are guaranteed to be bounded and deadlock free. Our case studies are video decoders for high definition video streaming such as VC-1 or HEVC.

This research is the central topic of Vagelis Bebelis' PhD thesis. It is conducted in collaboration with STMicroelectronics.

6.5. Static Analysis and Abstract Interpretation

Participants: Alain Girault, Bertrand Jeannet [contact person], Peter Schrammel.

6.5.1. Translating data-flow languages for hybrid systems simulation to hybrid automata for hybrid systems verification

Hybrid systems are used to model embedded computing systems interacting with their physical environment. There is a conceptual mismatch between high-level hybrid system languages like SIMULINK[34], which are used for simulation, and hybrid automata, the most suitable representation for safety verification. Indeed, in simulation languages the interaction between discrete and continuous execution steps is specified using the concept of zero-crossings, whereas hybrid automata exploit the notion of staying conditions.

In the context of the INRIA large scale action SYNCHRONICS (see §8.1.1.1), we studied how to translate the ZELUS hydrid data-flow language [34] developped in this project into logico-numerical hybrid automata by carefully pointing out this issue. We investigated various zero-crossing semantics, proposed a sound translation, and discussed to which extent the original semantics is preserved.

This work is part of the PhD thesis of Peter Schrammel and was presented at the conference HSCC'2012 (Hybrid Systems: Computation and Control) [22], [27].

6.5.2. Abstract Acceleration of general linear loops

We investigated abstract acceleration techniques for computing loop invariants for linear loops with linear assignments in their body and guards defined by the conjunction of linear inequalities.

While standard abstract interpretation considers over approximations over the set of reachable states at any loop iteration, and relies on extrapolation (*a.k.a.* widening) for making the analysis converge, abstract acceleration captures the effect of the loop with a single, non-iterative transfer function applied to the reachable states at the loop head. The concept of abstract acceleration has already been applied to restricted form of linear loops, by us [16] and others [58], and extended to logico-numerical loops [16]; the novelty here is to investigate general linear loops.

[34] http://www.mathworks.com

The main idea we developped is to over-approximate the set of transformation matrices associated to any number of iterations of the loop body and to apply this "accelerated" transformation to the initial states. This over-approximation is based on the Jordan normal form decomposition that allows deriving closed form symbolic expressions for the entries of the matrix modeling the effect of exactly n iterations of the loop. We then discover linear relationships between the symbolic expressions that hold for any number of iterations, and we obtain a set of matrices described by a polyhedra on its coefficients, which can be applied to a set of vectors also described by a convex polyhedra.

We also developed a technique to take into account the guard of the loop by bounding the number of loop iterations, which relies again on the Jordan normal form decomposition.

These ideas were implemented and evaluated on a series of simple loops, alone or inside outer loops, exhibiting classical behaviors: polynomial, stable and unstable exponential, inward spirals (damped oscillators), ... Our approach enables proofs that are out of the reach of most other techniques, that are either too unprecise (classical abstract interpretation) or limited to a restricted class of loops, *e.g.*, translation with resets in the case of abstract acceleration, or stable loops (in the sense of control theory) for ellipsoid methods.

This work was initiated during a visit to the University of Colorado-Boulder and is under review.

6.5.3. Logico-Numerical Max-Strategy Iteration

Strategy iteration methods aim at solving fixed point equations and are an alternative to abstract acceleration for fighting against the loss of precision incurred by extrapolation in classical interpretation. It has been shown that they improve precision in static analysis based on abstract interpretation and template abstract domains, *e.g.*, intervals, octagons or template polyhedra. However, they are limited to numerical programs.

We investigated a method for applying max-strategy iteration to logico-numerical programs, that is, programs with numerical and Boolean variables, without explicitly enumerating the Boolean state space. The method is optimal in the sense that it computes the *least fixed pointw.r.t.* the abstract domain.

Our experiments showed that the resulting logico-numerical max-strategy iteration gains one order of magnitude in terms of efficiency in comparison to the purely numerical approach while being almost as precise. Moreover, they are the first experimental results of applying max-strategy iteration to larger programs. This work has been accepted at VMCAI'2013 [23].

6.6. Component-Based Construction

Participants: Alain Girault, Gregor Goessler [contact person], Roopak Sinha, Gideon Smeding.

6.6.1. Incremental converter synthesis

We have proposed and implemented a formal incremental converter-generation algorithm for system-on-chip (SoC) designs. The approach generates a converter, if one exists, to control the interaction between multiple intellectual property (IP) protocols with possible control and data mismatches, and allows pre-converted systems to be re-converted with additional IPs in the future. IP protocols are represented using labeled transition systems (LTS), a simple but elegant abstraction framework which can be extracted from and converted to standard IP description languages such as VHDL. The user can provide control properties, each stated as an LTS with accepting states, to describe desired aspects of the converted system, including fairness and liveness. Furthermore, data specifications can be provided to bound data channels between interacting IPs such that they do not over/under flow. The approach takes into account the uncontrollable environment of a system by allowing users to identify signals exchanged between the SoC and the environment, which the converter can neither suppress nor generate.

Given these inputs, the conversion algorithm first computes the reachable state-space of a maximal non-deterministic converter that ensures (i) the satisfaction of the given data specifications and (ii) the trace equivalence with the given control specifications, using a greatest fix-point computation. It then checks, using the standard algorithm for Büchi games, whether the converter can ensure the satisfaction of the given control specifications (reachability of accepting states) regardless of how the environment behaves. If this is found to be true, deterministic converters can be automatically generated from the maximal non-deterministic converter generated during the first step. The algorithm is proven to be sound and complete, with a polynomial complexity in the state-space sizes of given IP protocols and specifications. It is also shown that it can be used for incremental design of SoCs, where IPs and specifications are added to an SoC in steps. Incremental design allows to constrain the combinatorial explosion of the explored state-space in each step, and also reduces on-chip wire congestion by decentralizing the conversion process.

A Java implementation has been created, and experimental results show that the algorithm can handle complex IP mismatches and specifications in medium to large AMBA based SoC systems. Future work involves creating a library of commonly-encountered specifications in SoC design such as sharing of control signals between interacting IPs using buffers, signal lifespans, and the generation of optimal converters based on quantitative criteria such as minimal power usage.

This work has been done within the AFMES associated team with the Electric and Computer Engineering Department of the University of Auckland.

6.6.2. Analysis of logical causality

The failure of one component may entail a cascade of failures in other components; several components may also fail independently. In such cases, elucidating the exact scenario that led to the failure is a complex and tedious task that requires significant expertise.

The notion of causality *(did an event e cause an event e'?)* has been studied in many disciplines, including philosophy, logic, statistics, and law. The definitions of causality studied in these disciplines usually amount to variants of the counterfactual test "e is a cause of e' if both e and e' have occurred, and in a world that is as close as possible to the actual world but where e does not occur, e' does not occur either". Surprisingly, the study of logical causality has so far received little attention in computer science, with the notable exception of [62] and its instantiations. However, this approach relies on a causal model that may not be known, for instance in presence of black-box components.

In [6] we have proposed a formal framework for reasoning about causality, based on black-box components interacting according to well identified *interaction models* [5].

We are currently extending to framework to other models of computation and communication, in particular, to timed automata, and developing a refinement of our original approach that reduces the number of false positives.

6.6.3. A Theory of fault recovery for component-based models

In [35][18] we have introduced a theory of fault recovery for component-based models. A model is specified in terms of a set of atomic components that are incrementally composed and synchronized by a set of glue operators. We have defined what it means for such models to provide a recovery mechanism, so that the model converges to its normal behavior in the presence of faults (*e.g.*, in self-stabilizing systems). We have identified corrector components whose presence in a model is essential to guarantee recovery after the occurrence of faults. We have also formalized component based models that effectively separate recovery from functional concerns. We have shown that any model that provides fault recovery can be transformed into an equivalent model, where functional and recovery tasks are modularized in different components.

6.7. Aspect-Oriented Programming

Participants: Dmitry Burlyaev, Pascal Fradet [contact person], Alain Girault.

The goal of Aspect-Oriented Programming (AOP) is to isolate aspects (such as security, synchronization, or error handling) which cross-cut the program basic functionality and whose implementation usually yields tangled code. In AOP, such aspects are specified separately and integrated into the program by an automatic transformation process called *weaving*.

Although this paradigm has great practical potential, it still lacks formalization and undisciplined uses make reasoning on programs very difficult. Our work on AOP addresses these issues by studying foundational issues (semantics, analysis, verification) and by considering domain-specific aspects (availability, fault tolerance or refinement aspects) as formal properties.

6.7.1. Aspects preserving properties

Aspect Oriented Programming can arbitrarily distort the semantics of programs. In particular, weaving can invalidate crucial safety and liveness properties of the base program.

We have identified categories of aspects that preserve some classes of properties [13]. Our categories of aspects comprise, among others, observers, aborters, and confiners. For example, observers do not modify the base program's state and control-flow (*e.g.*, persistence, profiling, and debugging aspects). These categories are defined formally based on a language independent abstract semantic framework. The classes of properties are defined as subsets of LTL for deterministic programs and CTL* for non-deterministic ones. We have formally proved that, for any program, the weaving of any aspect in a category preserves any property in the related class. In a second step, we have designed for each aspect category a specialized aspect language which ensures that any aspect written in that language belongs to the corresponding category. These languages preserve the corresponding classes of properties by construction.

This work was conducted in collaboration with Rémi Douence from the ASCOLAINRIA team at École des Mines de Nantes.

6.7.2. Fault tolerance aspects

In the recent years, we have studied the implementation of specific fault tolerance techniques in real-time embedded systems using program transformation [1]. We are now investigating the use of fault-tolerance aspects in digital circuits. To this aim, we consider program transformations for hardware description languages (HDL). Our goal is to design an aspect language allowing users to specify and tune a wide range of fault tolerance techniques, while ensuring that the woven HDL program remains synthesizable. The advantage would be to produce fault-tolerant circuits by specifying fault-tolerant strategies separately from the functional specifications.

We have reviewed the different fault tolerant techniques used in integrated circuits: concurrent error detection, error detecting and correcting codes (Hamming, Berger codes, ...), spatial and time redundancy. We have designed a simple hardware description language inspired from LUSTRE and Lucid Synchrone. It is a core functional language manipulating synchronous boolean streams. Faults are represented by bit flips and we take into account all fault models of the form “at most 1 faults within n clock signals”. The language semantics as well as the fault model have been formalized in Coq. Many basic (library) properties have been shown on that language.

We are currently expressing different variants of triple modular redundancy (TMR) as program transformations. We are also studying optimizations to prevent the insertion of useless voters in TMR. The next step is to prove that these transformations make the programs fault tolerant *w.r.t.* specific fault models. Further work also includes the study of mixed techniques (*e.g.*, spatial and time redundancy), their high level specification using an AOP-like language and their implementation as transformations.

PROSECCO Project-Team

6. New Results

6.1. Verification of Security Protocols in the Symbolic Model

The symbolic model of protocols, or Dolev-Yao model is an abstract model in which messages are represented by terms. Our protocol verifier PROVERIF relies on this model. This year, we have mainly worked on the verification of protocols with lists and on an extension of PROVERIF to prove more observational equivalences.

6.1.1. Verification of Protocols with Lists

Participants: Bruno Blanchet [correspondant], Miriam Paiola.

security protocols, symbolic model, automatic verification, Horn clauses, secrecy

We have designed a novel, simple technique for proving secrecy properties for security protocols that manipulate lists of unbounded length, for an unbounded number of sessions [32]. More specifically, our technique relies on the Horn clause approach used in the automatic verifier PROVERIF: we show that if a protocol is proven secure by our technique with lists of length one, then it is secure for lists of unbounded length. Interestingly, this theorem relies on approximations made by our verification technique: in general, secrecy for lists of length one does not imply secrecy for lists of unbounded length. Our result can be used in particular to prove secrecy properties for group protocols with an unbounded number of participants and for some XML protocols (web services) with PROVERIF.

6.1.2. Proving More Process Equivalences with ProVerif

Participants: Bruno Blanchet [correspondant], Vincent Cheval.

security protocols, symbolic model, automatic verification, observational equivalence, privacy

We have extended the automatic protocol verifier PROVERIF in order to prove more observational equivalences [28]. PROVERIF can prove observational equivalence between processes that have the same structure but differ by the messages they contain. In order to extend the class of equivalences that PROVERIF handles, we extend the language of terms by defining more functions (destructors) by rewrite rules. In particular, we allow rewrite rules with inequalities as side-conditions, so that we can express tests "if then else" inside terms. Finally, we provide an automatic procedure that translates a process into an equivalent process that performs as many actions as possible inside terms, to allow PROVERIF to prove the desired equivalence. These extensions have been implemented in PROVERIF and allow us to automatically prove anonymity in the private authentication protocol by Abadi and Fournet.

6.2. Verification of Security Protocols in the Computational Model

The computational model of protocols considers messages as bitstrings, which is more realistic than the formal model, but also makes the proofs more difficult. Our verifier CRYPTOVERIF is sound in this model. This year, we have worked on a compiler from CRYPTOVERIF speficications to OCaml, and we have used CRYPTOVERIF to verify the password-based protocol One-Encryption Key Exchange (OEKE).

6.2.1. Generation of Implementations Proved Secure in the Computational model

Participants: Bruno Blanchet [correspondant], David Cadé.

security protocols, computational model, implementation, verification, compiler

We have designed a novel approach for proving specifications of security protocols in the computational model and generating runnable implementations from such proved specifications. We rely on the computationally-sound protocol verifier CryptoVerif for proving the specification, and we have implemented a compiler that translates a CryptoVerif specification into an implementation in OCaml [26]. We have also proved that this compiler preserves security [27]. We have applied this compiler to the SSH Transport Layer protocol: we proved the authentication of the server and the secrecy of the session keys in this protocol and verified that the generated implementation successfully interacts with OpenSSH. The secrecy of messages sent over the SSH tunnel cannot be proved due to known weaknesses in SSH with CBC-mode encryption.

6.2.2. Proof of One-Encryption Key Exchange using CryptoVerif

Participant: Bruno Blanchet [correspondant].

security protocols, computational model, automatic proofs, formal methods, password-based authentication

We have obtained a mechanized proof of the password-based protocol One-Encryption Key Exchange (OEKE) using the computationally-sound protocol prover CryptoVerif [25]. OEKE is a non-trivial protocol, and thus mechanizing its proof provides additional confidence that it is correct. This case study was also an opportunity to implement several important extensions of CryptoVerif, useful for proving many other protocols. We have indeed extended CryptoVerif to support the computational Diffie-Hellman assumption. We have also added support for proofs that rely on Shoup's lemma and additional game transformations. In particular, it is now possible to insert case distinctions manually and to merge cases that no longer need to be distinguished. Eventually, some improvements have been added on the computation of the probability bounds for attacks, providing better reductions. In particular, we improve over the standard computation of probabilities when Shoup's lemma is used, which allows us to improve the bound given in a previous manual proof of OEKE, and to show that the adversary can test at most one password per session of the protocol.

6.3. New Attacks on RSA PKCS#1 v1.5

Participants: Graham Steel [correspondant], Romain Bardou.

cryptographic hardware, security API, key management, vulnerabilities

RSA PKCS#1v1.5 is the most commonly used standard for public key encryption, used for example in TLS/SSL. It has been known to be vulnerable to a so-called padding-oracle attack since 1998 when Bleichenbacher described the vulnerability at CRYPTO. The attack, known as the “million message attack” was not thought to present a practical threat, due in part to the large number of oracle messages required. In a paper published at CRYPTO 2012 [22] we gave original modifications showing how the attack can be completed in a median of just 15 000 messages. The results lead to widespread interest, indicated by over 1400 downloads of the long version of the paper from the HAL webpage and articles in the New York Times, Boston Globe and Süddeutscher Zeitung.

6.4. Security Proofs for Revocation

Participants: Graham Steel [correspondant], Véronique Cortier, Cyrille Wiedling.

security API, key management, formal methods, security proofs

Revocation of expired or corrupted keys is a common feature of industrially deployed key management systems but an aspect that is almost always missing from formal models. We succeeded in adding revocation to a formal specification of a key management API allowing the proof of strong security properties after corrupted keys are revoked. In particular we showed a self-healing property whereby after a corrupted key expires, after a certain amount of time, the system is safe again. The work was published at ACM CCS 2012.

6.5. Discovering Concrete Attacks on Web Applications by Formal Analysis

Participants: Karthikeyan Bhargavan [correspondant], Sergio Maffeis, Chetan Bansal, Antoine Delignat-Lavaud.

web application security, formal methods, automated verification, vulnerabilities Social sign-on and social sharing are becoming an ever more popular feature of web applications. This success is largely due to the APIs and support offered by prominent social networks, such as Facebook, Twitter, and Google, on the basis of new open standards such as the OAuth 2.0 authorization protocol. A formal analysis of these protocols must account for malicious websites and common web application vulnerabilities, such as cross-site request forgery and open redirectors. We model several configurations of the OAuth 2.0 protocol in the applied pi-calculus and verify them using ProVerif. Our models rely on WebSpi, a new library for modeling web applications and web-based attackers that is designed to help discover concrete website attacks. Our approach is validated by finding dozens of previously unknown vulnerabilities in popular websites such as Yahoo and WordPress, when they connect to social networks such as Twitter and Facebook. This work was published in CSF'12 [21].

To protect sensitive user data against server-side attacks, a number of security-conscious web applications have turned to client-side encryption, where only encrypted user data is ever stored in the cloud. We formally investigate the security of a number of such applications, including password managers, cloud storage providers, an e-voting website and a conference management system. We show that their security relies on both their use of cryptography and the way it combines with common web security mechanisms as implemented in the browser. We model these applications using the WebSpi web security library for ProVerif, we discuss novel attacks found by automated formal analysis, and we propose robust countermeasures. Some of the attacks we discovered were presented at WOOT'12 [24]. Our formal models and verified countermeasures are going to be presented at POST'13 [20].

6.6. Attacks and Proofs for TLS Implementations

Participants: Alfredo Pironti [correspondant], Karthikeyan Bhargavan, Pierre-Yves Strub, Cedric Fournet, Markulf Kohlweiss.

cryptographic protocol, formal methods, automated verification, traffic analysis, vulnerabilities

TLS is possibly the most used secure communications protocol, with a 18-year history of flaws and fixes, ranging from its protocol logic to its cryptographic design, and from the Internet standard to its diverse implementations. We have been engaged in a long-term project on verifying TLS implementations and this project is now coming to fruition, with a number of papers are now in the pipeline. We list two new results below, both are submitted for review.

Websites commonly use HTTPS to protect their users' private data from network-based attackers. By combining public social network profiles with TLS traffic analysis, we present a new attack that reveals the precise identities of users accessing major websites. As a countermeasure, we propose a novel length-hiding scheme that leverages standard TLS padding to enforce website-specific privacy policies. We present several implementations of this scheme, notably a patch for GnuTLS that offers a rich length-hiding API and an Apache module that uses this API to enforce an anonymity policy for sensitive user files. Our implementations are the first to fully exercise the length-hiding features of TLS and our work uncovers hidden timing assumptions in recent formal proofs of these features. Compared to previous work, we offer the first countermeasure that is standards-based, provably secure, and experimentally effective, yet pragmatic, offering websites a precise trade-off between user privacy and bandwidth efficiency. This work is available as an Inria technical report [36].

We develop a verified reference implementation of TLS 1.2. Our code fully supports its wire formats, ciphersuites, sessions and connections, re-handshakes and resumptions, alerts and errors, and data fragmentation, as prescribed in the RFCs; it interoperates with mainstream web browsers and servers. At the same time, our code is carefully structured to enable its modular, automated verification, from its main API down to computational assumptions on its cryptographic algorithms. Our implementation is written in F# and specified in F7. We present security specifications for its main components, such as authenticated stream encryption for the record layer and key establishment for the handshake. We describe their verification using the F7 refinement typechecker. To this end, we equip each cryptographic primitive and construction of TLS with a new typed interface that captures its security prop- erties, and we gradually replace concrete implementations with ideal functionalities. We finally typecheck the protocol state machine, and thus obtain precise security theorems for TLS, as it is implemented and deployed. We also revisit classic attacks and report a few new ones. This work is under review and will be released as an Inria technical report in January 2013.

S4 Project-Team

6. New Results

6.1. Petri Nets and their Synthesis

Participants: Eric Badouel, Philippe Darondeau.

6.1.1. Deciding Selective Declassification of Petri Nets

In [20], we consider declassification, as effected by downgrading actions D, in the context of intransitive non-interference encountered in systems that consist of high-level (secret) actions H and low-level (public) actions L. In a previous work, we had shown the decidability of a strong form of declassification, by which D contains only a single action d declassifying all H actions at once. We continue this study by considering selective declassification, where each transition d in D can declassify a subset $H(d)$ of H. The decidability of this more flexible, application-prone declassification framework is proved in the context of (possibly unbounded) Petri nets with possibly infinite state spaces.

6.1.2. Petri Net Distributability

A Petri net is distributed if, given an allocation of transitions to (geographical) locations, no two transitions at different locations share a common input place. A system is distributable if there is some distributed Petri net implementing it. We address in [21] the question of which systems can be distributed, while respecting a given allocation. We state the problem formally and discuss several examples illuminating — to the best of our knowledge — the current status of this work.

6.1.3. Petri Net Reachability Graphs: Decidability Status of First Order Prioperties

We investigated in [13] the decidability and complexity status of model-checking problems on unlabelled reachability graphs of Petri nets by considering first-order, modal and pattern-based languages without labels on transitions or atomic propositions on markings. We have considered several parameters to separate decidable problems from undecidable ones. Not only were we able to provide precise borders and a systematic analysis, but we also demonstrated the robustness of our proof techniques.

6.1.4. α-reconstructibility of Workflow Nets

The α-algorithm is a process mining algorithm, introduced by van der Aalst et al, that constructs a workflow net from an event log. A class of nets, the structured workflow nets, was recognized to be reconstructible by algorithm α from their language (or a representative subset of it). In [14] we assessed more precisely the α-algorithm we provided a characterization of the class of the workflow nets that are discovered by α.

6.2. Hybrid Modeling

Participants: Albert Benveniste, Benoît Caillaud.

Hybrid system modelers have become a corner stone of complex embedded system development. Embedded systems include not only control components and software, but also physical devices. In this area, Simulink is a de facto standard design framework, and Modelica a new player. However, such tools raise several issues related to the lack of reproducibility of simulations (sensitivity to simulation parameters and to the choice of a simulation engine). In [10] we propose using techniques from non-standard analysis to define a semantic domain for hybrid systems. Non-standard analysis is an extension of classical analysis in which infinitesimal (the ϵ and η in the celebrated generic sentence $\forall\epsilon\exists\eta$... of college maths) can be manipulated as first class citizens. This approach allows us to define both a denotational semantics, a constructive semantics, and a Kahn Process Network semantics for hybrid systems, thus establishing simulation engines on a sound but flexible mathematical foundation. These semantics offer a clear distinction between the concerns of the numerical analyst (solving differential equations) and those of the computer scientist (generating execution schemes). We also discuss a number of practical and fundamental issues in hybrid system modelers that give rise to non-reproducibility of results, non-determinism, and undesirable side effects. Of particular importance are cascaded mode changes (also called "zero-crossings" in the context of hybrid systems modelers). This work has taken place in the framework of the Synchronics large scale initiative (see section 7.1.1).

6.3. Component-Based Design

Participants: Albert Benveniste, Benoît Caillaud, Sophie Pinchinat.

6.3.1. Application of Interface Theories to the Separate Compilation of Synchronous Programs

We study in [15], [26] the problem of separate compilation, i.e., the generation of modular code, for the discrete time part of block-diagrams formalisms such as Simulink, Modelica, or Scade. Code is modular in that it is generated for a given composite block independently from context (i.e., without knowing in which diagrams the block is to be used) and using minimal information about the internals of the block. Just using off-the-shelf C code generation (e.g., as available in Simulink) does not provide modular code. Separate compilation was solved by Lublinerman et al. for the special case of single clocked diagrams, in which all signals are updated at a same unique clock. For the same case, Pouzet and Raymond proposed algorithms that scale-up properly to real-size applications. The technique of Lublinerman et al. was extended to some classes of multi-clocked and timed diagrams. We study this problem in its full generality and we show that it can be cast to a special class of controller synthesis problems by relying on recently proposed modal interface theories.

6.3.2. Contracts for System Design

Systems design has become a key challenge and differentiating factor over the last decades for system companies. Aircrafts, trains, cars, plants, distributed telecommunication military or health care systems, and more, involve systems design as a critical step. Complexity has caused system design times and costs to go severely over budget so as to threaten the health of entire industrial sectors. Heuristic methods and standard practices do not seem to scale with complexity so that novel design methods and tools based on a strong theoretical foundation are sorely needed. Model-based design as well as other methodologies such as layered and compositional design have been used recently but a unified intellectual framework with a complete design flow supported by formal tools is still lacking albeit some attempts at this framework such as Platform-based Design have been successfully deployed. Recently an “orthogonal” approach has been proposed that can be applied to all methodologies proposed thus far to provide a rigorous scaffolding for verification, analysis and abstraction/refinement: contract-based design. Several results have been obtained in this domain but a unified treatment of the topic that can help in putting contract-based design in perspective is still missing. In [25], we intend to provide such treatment where contracts are precisely defined and characterized so that they can be used in design methodologies such as the ones mentioned above with no ambiguity. In addition, the paper provides an important link between interfaces and contracts to show similarities and correspondences. Examples of the use of contracts in system design are provided, including one based on Modal Interfaces, using the Mica tool (see section 5.1).

6.3.3. Ensuring Reachability by Design

In [18], [28], we study the independent implementability of reachability properties, which are in general not compositional. We consider modal specifications, which are widely acknowledged as suitable for abstracting implementation details of components while exposing to the environment relevant information about cross-component interactions. In order to obtain the required expressivity, we extend them with marked states to model states to be reached. We then develop an algebra with both logical and structural composition operators ensuring reachability properties by construction.

6.3.4. Modal event-clock specifications for timed component-based design

Modal specifications are classic, convenient, and expressive mathematical objects to represent interfaces of component-based systems. However, time is a crucial aspect of systems for practical applications, e.g. in the area of embedded systems. And yet, only few results exist on the design of timed component-based systems. In [11], we propose a timed extension of modal specifications, together with fundamental operations (conjunction, product, and quotient) that enable reasoning in a compositional way about timed system. The specifications are given as modal event-clock automata, where clock resets are easy to handle. We develop an entire theory that promotes efficient incremental design techniques.

6.4. Automata, Games and Logics for Supervisory Control and System Synthesis

Participants: Philippe Darondeau, Bastien Maubert, Sophie Pinchinat.

6.4.1. Distributed Control of Discrete Event Systems: A First Step

To initiate a discussion on the modeling requirements for distributed control of discrete-event systems, a partially-automated region-based methodology is presented in [23]. The methodology is illustrated via a well-known example from distributed computing: the dining philosophers.

6.4.2. Enforcing Opacity of Regular Predicates on Modal Transition Systems

In [22] we considered the following problem: Given a labelled transition system LTS partially observed by an attacker, and a regular predicate Sec over the runs of LTS, enforcing opacity of the secret Sec in LTS means computing a supervisory controller K such that an attacker who observes a run of K/LTS cannot ascertain that the trace of this run belongs to Sec based on the knowledge of LTS and K. We then lifted the problem from a single labelled transition system LTS to the class of all labelled transition systems specified by a modal transition system MTS. The lifted problem is to compute the maximally permissive controller K such that Sec is opaque in K/LTS for every labelled transition system LTS which is a model of MTS. The situations of the attacker and of the controller are dissymmetric: at run time, the attacker may fully know LTS and K whereas the controller knows only MTS and the sequence of actions executed so far by the unknown LTS. We addressed the problem in two cases. Let Σ_a denote the set of actions that can be observed by the attacker, and let Σ_c and Σ_o denote the sets of actions that can be controlled and observed by the controller, respectively. We provided optimal and regular controllers that enforce the opacity of regular secrets when $\Sigma_c \subseteq \Sigma_o \subseteq \Sigma_a = \Sigma$. We also provided optimal and regular controllers that enforce the opacity of regular upper-closed secrets ($Sec = Sec.\Sigma^*$) when $\Sigma_a \subseteq \Sigma_c \subseteq \Sigma_o = \Sigma$.

6.4.3. Analysis of partially observed recursive tile systems

The analysis of discrete event systems under partial observation is an important topic, with major applications such as the detection of information flow and the diagnosis of faulty behaviors. In [19], we consider recursive tile systems, which are infinite systems generated by a finite collection of finite *tiles*, a simplified variant of deterministic graph grammars. Recursive tile systems are expressive enough to capture classical models of recursive systems, such as the pushdown systems and the recursive state machines. They are infinite-state in general and therefore standard powerset constructions for monitoring do not always apply. We exhibit computable conditions on recursive tile systems and present non-trivial constructions that yield effective computation of the monitors. We apply these results to the classic problems of opacity and diagnosability.

6.4.4. Uniform Strategies

In [29], we consider turn-based game arenas for which we investigate uniformity properties of strategies. These properties involve bundles of plays, that arise from some semantical motive. Typically, we can represent constraints on allowed strategies, such as being observation-based. We propose a formal language to specify uniformity properties and demonstrate its relevance by rephrasing various known problems from the literature. Note that the ability to correlate different plays cannot be achieved by any branching-time logic if not equipped with an additional modality, so-called R in this contribution. We also study an automated procedure to synthesize strategies subject to a uniformity property, which strictly extends exitsting results based on, say standard temporal logics. We exhibit a generic solution for the synthesis problem provided the bundles of plays rely on any binary relation definable by a finite state transducer. This solution yields a non-elementary procedure.

6.4.5. Emptiness Of Alternating Parity Tree Automata Using Games With Imperfect Information

In [30], we focus on the emptiness problem for alternating parity tree automata. The usual technique to tackle this problem first removes alternation, going to non-determinism, and then checks emptiness by reduction to a two-player perfect-information parity game. In this note, we give an alternative roadmap to this problem by providing a direct reduction to the emptiness problem to solving an imperfect-information two-player parity game.

6.4.6. On timed alternating simulation for concurrent timed games

We address in [12] the problem of alternating simulation refinement for concurrent timed games (*TG*). We show that checking timed alternating simulation between *TG* is EXPTIME-complete, and provide a logical characterization of this preorder in terms of a meaningful fragment of a new logic, *TAMTL**. *TAMTL** is an action-based timed extension of standard alternating-time temporal logic *ATL**, which allows to quantify over strategies where the designated coalition of players is not responsible for blocking time. While for full *TAMTL**, model-checking *TG* is undecidable, we show that for its fragment *TAMTL*, corresponding to the timed version of *ATL*, the problem is instead decidable and in EXPTIME.

SECRET Project-Team

5. New Results

5.1. Symmetric cryptosystems

Participants: Christina Boura, Baudoin Collard, Anne Canteaut, Pascale Charpin, Gohar Kyureghyan, María Naya-Plasencia, Joëlle Roué, Valentin Suder.

From outside, it might appear that symmetric techniques become obsolete after the invention of public-key cryptography in the mid 1970's. However, they are still widely used because they are the only ones that can achieve some major features as high-speed or low-cost encryption, fast authentication, and efficient hashing. Today, we find symmetric algorithms in GSM mobile phones, in credit cards, in WLAN connections. Symmetric cryptology is a very active research area which is stimulated by a pressing industrial demand for low-cost implementations (in terms of power consumption, gate complexity...). These extremely restricting implementation requirements are crucial when designing secure symmetric primitives and they might be at the origin of some weaknesses. Actually, these constraints seem quite incompatible with the rather complex mathematical tools needed for constructing a provably secure system.

The specificity of our research work is that it considers all aspects of the field, from the practical ones (new attacks, concrete specifications of new systems) to the most theoretical ones (study of the algebraic structure of underlying mathematical objects, definition of optimal objects). But, our purpose is to study these aspects not separately but as several sides of the same domain. Our approach mainly relies on the idea that, in order to guarantee a provable resistance to the known attacks and to achieve extremely good performance, a symmetric cipher must use very particular building blocks, whose algebraic structures may introduce unintended weaknesses. Our research work captures this conflict for all families of symmetric ciphers. It includes new attacks and the search for new building blocks which ensure both a high resistance to the known attacks and a low implementation cost. This work, which combines cryptanalysis and the theoretical study of discrete mathematical objects, is essential to progress in the formal analysis of the security of symmetric systems.

In this context, the very important challenges are the designs of low-cost ciphers and of secure hash functions. Most teams in the research community are actually working on the design and on the analysis (cryptanalysis and optimisation of the performance) of such primitives.

5.1.1. Hash functions.

Following the recent attacks against almost all existing hash functions (MD5, SHA-0, SHA-1...), we have initiated a research work in this area, especially within the Saphir-2 ANR project and with several PhD theses. Our work on hash functions is two-fold: we have designed two new hash functions, named FSB and Shabal, which have been submitted to the SHA-3 competition, and we have investigated the security of several hash functions, including the previous standards (SHA-0, SHA-1...) and some other SHA-3 candidates.

Recent results:

- Upper bounds on the degree of an iterated permutation from the degree of the inverse of the inner transformation; this result has been applied both to hash functions and to block ciphers. Most notably, this work leads to the best (theoretical) analysis of the hash function Keccak, which has been selected for the new SHA-3 standard [12], [22], [9].
- Side-channel attacks on two SHA-3 candidates, Skein and Grøstl, when they are used with HMAC, and counter-measures [23], [50].
- Indifferentiability results for a broadened mode of operation including the modes based on block ciphers, and modes based on un-keyed functions [51].

5.1.2. Block ciphers.

Even if the security of the current block cipher standard, AES, is not threaten when it is used in a classical context, there is still a need for the design of improved attacks, and for the determination of design criteria which guarantee that the existing attacks do not apply. This notably requires a deep understanding of all previously proposed attacks. Moreover, there is a high demand from the industry of lightweight block ciphers for some constrained environments. Several such algorithms have been proposed in the last few years and their security should be carefully analysed. Most of our work in this area is related to an ANR Project named BLOC.

Recent results:

- Algebraic analysis of some recent lightweight block ciphers, including LED and Piccolo [24].
- Analysis of the security of the lightweight block cipher mCRYPTON [56].
- Design of a new block cipher, named PRINCE, with a very low-latency, leading to instantaneous encryption (i.e., within one clock cycle) with a very competitive chip area [21], [49].
- Analysis of the differential properties of the AES Superbox [58].
- Study of the significance of the related-key and known-key models for block ciphers [48].

5.1.3. Stream ciphers.

The project-team has been involved in the international project eSTREAM, which aimed at recommending some secure stream ciphers.

Recent results:

- Generalisation of several improvements of the so-called correlation attacks against stream ciphers and study of their complexities [13].
- Study of the bias of parity-check relations for combination generators used in stream ciphers [14].

5.1.4. Cryptographic properties and construction of appropriate building blocks.

The construction of building blocks which guarantee a high resistance to the known attacks is a major topic within our project-team, for stream ciphers, block ciphers and hash functions. The use of such optimal objects actually leads to some mathematical structures which may be the origin of new attacks. This work involves fundamental aspects related to discrete mathematics, cryptanalysis and implementation aspects. Actually, characterising the structures of the building blocks which are optimal regarding to some attacks is very important for finding appropriate constructions and also for determining whether the underlying structure induces some weaknesses or not.

For these reasons, we have investigated several families of filtering functions and of S-boxes which are well-suited for their cryptographic properties or for their implementation characteristics. For instance, bent functions, which are the Boolean functions which achieve the highest possible nonlinearity, have been extensively studied in order to provide some elements for a classification, or to adapt these functions to practical cryptographic constructions. We have also been interested in functions with a low differential uniformity (*e.g.*, APN functions), which are the S-boxes ensuring an (almost) optimal resistance to differential cryptanalysis.

Recent results:

- Study of the algebraic properties (e.g. the algebraic degree) of the inverses of APN power permutations [26].
- Study of the planarity of some mappings, including products of linearized polynomials [25], [16].
- Definition of a new criterion for Sboxes and link with some recent algebraic attacks on the hash functuion Hamsi [29], [9].
- Survey of PN and APN mappings [42].

5.2. Code-based cryptography

Participants: Grégory Landais, Rafael Misoczki, Nicolas Sendrier, Dimitrios Simos, Jean-Pierre Tillich.

Most popular public-key cryptographic schemes rely either on the factorization problem (RSA, Rabin), or on the discrete logarithm problem (Diffie-Hellman, El Gamal, DSA). These systems have evolved and today instead of the classical groups ($\mathbf{Z}/n\mathbf{Z}$) we may use groups on elliptic curves. They allow a shorter block and key size for the same level of security. An intensive effort of the research community has been and is still being conducted to investigate the main aspects of these systems: implementation, theoretical and practical security. It must be noted that these systems all rely on algorithmic number theory. As they are used in most, if not all, applications of public-key cryptography today (and it will probably remain so in the near future), cryptographic applications are thus vulnerable to a single breakthrough in algorithmics or in hardware (a quantum computer can break all those scheme).

Diversity is a way to dilute that risk, and it is the duty of the cryptographic research community to prepare and propose alternatives to the number theoretic based systems. The most serious tracks today are lattice-based cryptography (NTRU,...), multivariate cryptography (HFE,...) and code-based cryptography (McEliece encryption scheme,...). All these alternatives are referred to as *post-quantum cryptosystems*, since they rely on difficult algorithmic problems which would not be solved by the coming-up of the quantum computer.

The code-based primitives have been investigated in details within the project-team. The first cryptosystem based on error-correcting codes was a public-key encryption scheme proposed by McEliece in 1978; a dual variant was proposed in 1986 by Niederreiter. We proposed the first (and only) digital signature scheme in 2001. Those systems enjoy very interesting features (fast encryption/decryption, short signature, good security reduction) but also have their drawbacks (large public key, encryption overhead, expensive signature generation). Some of the main issues in this field are

- security analysis , implementation and practicality of existing solutions,
- reducing the key size, *e.g.*, by using rank metric instead of Hamming metric, or by using particular families of codes,
- address new functionalities, like hashing or symmetric encryption.

Recent results:

- A new variant of McEliece using Moderate Density Parity Check (MDPC) codes [55];
- An optimized software implementation of the code-based digital signature scheme CFS [27];
- An attack on a homomorphic encryption scheme [53];
- An attack on a variant of the McEliece cryptosystem based on Reed-Solomon codes [54].

5.3. Error-correcting codes and applications

Participants: Mamdouh Abbara, Marion Bellard, Denise Maurice, Nicolas Sendrier, Jean-Christophe Sibel, Jean-Pierre Tillich, Audrey Tixier.

We mainly investigate two new application domains for decoding algorithms: reverse engineering of communication systems, and quantum error correcting codes for which we have shown that some of them can be decoded successfully with iterative decoding algorithms.

5.3.1. Quantum codes.

The knowledge we have acquired in iterative decoding techniques has also led to study whether or not the very same techniques could also be used to decode quantum codes. Part of the old ACI project “RQ” in which we were involved and the new ANR project “COCQ” are about this topic. It is worth noticing that protecting quantum information from external noise is an issue of paramount importance for building a quantum computer. It also worthwhile to notice that all quantum error-correcting code schemes proposed up to now suffer from the very same problem that the first (classical) error-correcting codes had: there are constructions of good quantum codes, but for the best of them it is not known how to decode them in polynomial time. Our approach for overcoming this problem has been to study whether or not the family of turbo-codes and LDPC codes (and the associated iterative decoding algorithms) have a quantum counterpart.

Recent results:

- Construction of quantum LDPC codes obtained by transforming a quantum CSS LDPC code into a code over a larger alphabet which improves substantially the performances under iterative decoding [18];
- Construction of spatially coupled quantum LDPC codes which performs well under iterative decoding almost up to the coherent capacity of the quantum channel [19].

5.3.2. Reverse engineering of communication systems.

To evaluate the quality of a cryptographic algorithm, it is usually assumed that its specifications are public, as, in accordance with Kerckhoffs principle [1], it would be dangerous to rely, even partially, on the fact that the adversary does not know those specifications. However, this fundamental rule does not mean that the specifications are known to the attacker. In practice, before mounting a cryptanalysis, it is necessary to strip off the data. This reverse engineering process is often subtle, even when the data formatting is not concealed on purpose. A typical case is interception; some raw data, not necessarily encrypted, is observed out of a noisy channel. To access the information, the whole communication system has first to be disassembled and every constituent reconstructed. Our activity within this domain, whose first aim is to establish the scientific and technical foundations of a discipline which does not exist yet at an academic level, has been supported by some industrial contracts driven by the DGA and the French Ministry for Defense.

Recent results:

- Recontruction of the constellation labeling (i.e. used in the modulator of a communication system) in presence of error and when the underlying code is convolutional [20].

[1] Kerckhoffs stated that principle in a paper entitled *La Cryptographie militaire*, published in 1883.

SECSI Project-Team

6. New Results

6.1. Dishonest keys (Objective 2)

Participants: Hubert Comon-Lundh, Guillaume Scerri.

One of the main issues in the formal verification of the security protocols is the validity (and scope) of the formal model. Otherwise, it may happen that a protocol is proved and later someone finds an attack. This paradoxical situation may happen when the formal model used in the proof is too abstract.

A main stream of research therefore consists in proving full abstraction results (also called *soundness*): if the protocol is secure in the (symbolic) model, then an attack can only occur with negligible probability in a computational model. Such results have two main drawbacks: first they are very complicated, and have to be completed again and again for each combination of security primitives. Second, they require strong hypotheses on the primitives, some of which are not realistic. For instance, it is assumed that the attacker cannot forge his own keys (or that all keys come with their certificates, even for symmetric encryption keys).

Hubert Comon-Lundh, Véronique Cortier and Guillaume Scerri [31] propose an extension of the symbolic model, and prove it computationally sound, without this restriction on the dishonest keys.

6.2. Unconditional Soundness (Objective 2)

Participant: Hubert Comon-Lundh.

Hubert Comon-Lundh, Véronique Cortier and Guillaume Scerri [31] show how one can drop one of the assumptions of computational soundness results. However, the proofs remain very complicated and there are still assumptions such as the absence of key cycles, or no dynamic corruption... that are still necessary for all these results.

Gergei Bana and Hubert Comon-Lundh investigated a completely different approach to formal security proofs [25], which does not make any such assumptions. The idea can be stated in a nutshell: whereas all existing formal models state the attacker's abilities, they propose to formally state what the attacker *cannot* do.

This makes a big difference, since the soundness need only to be proved formula by formula and only the very necessary assumptions are used for such formulas (for instance, no absence of key cycles is needed). This does not need to be proved again when a primitive is added.

The counterpart of this nice approach is the difficulty of the automation: a tool is required for checking the consistency of a set of axioms, together with the conditions accumulated along a trace. This problem is the subject of research for the next year(s).

6.3. QRB-Domains (Objective 4)

Participant: Jean Goubault-Larrecq [correspondant].

One of the outstanding problems that remains in the denotational semantics of higher-order programming languages with probabilistic choice is the existence of a suitable, convenient category of domains for defining the denotations of types. Technically, a category of so-called continuous domains is sought after, which would be Cartesian-closed and stable by the action of the probabilistic powerdomain functor. This is not known to exist, and is part of the Jung-Tix conjecture. Jean Goubault-Larrecq found out that relaxing continuity to quasi-continuity helped gaining stability by the action of the probabilistic powerdomain functor [20]. This is an extended version of previous work published at the LICS'10 conference.

6.4. Complete WSTS

Participant: Jean Goubault-Larrecq [correspondant].

Well-structured transition systems form a large class of infinite-state transition systems on which one can decide coverability (a slightly relaxed form of reachability). These include Petri nets, lossy channel systems, and various process algebras.

With Alain Finkel, Jean Goubault-Larrecq developed a theory of *complete* well-structured transition systems, allowing one to generalize Karp and Miller's coverability tree construction for Petri nets to all well-structured transition systems. This work culminated in [19], following two conference papers (STACS'09, ICALP'09). The general theory was the topic of the invited talk [34].

6.5. Static Analysis of Programs with Imprecise Probabilities

Participant: Jean Goubault-Larrecq [correspondant].

Static analyses allows one to obtain guarantees about the behavior of programs, without running them. Programs that handle numerical data such as feedback control loops pose a challenge in this area. This gets even harder when one considers programs that read numerical data from sensors, and write to actuators, as these data are imprecise, and are governed by probability distributions that may themselves be unknown, and only know to fall into some interval of distributions. As part of the ANR projet blanc CPP, an efficient static analysis framework that deals with this kind of programs was proposed [16], based on P-boxes and Dempster-Shafer structures to handle imprecise probabilities. This is based on work first presented at the SCAN'11 conference.

6.6. New Attacks on RSA PKCS#1v1.5 (Objective 2)

Participants: Graham Steel [correspondant], Romain Bardou.

RSA PKCS#1v1.5 is the most commonly used standard for public key encryption, used for example in TLS/SSL. It has been known to be vulnerable to a so-called padding-oracle attack since 1998 when Bleichenbacher described the vulnerability at CRYPTO. The attack, known was the "million message attack" was not thought to present a practical threat, due in part to the large number of oracle messages required. In a paper published at CRYPTO 2012 [26] we gave original modifications showing how the attack can be completed in a median of just 15 000 messages. The results led to widespread interest, indicated by over 1400 downloads of the long version of the paper from the HAL webpage and articles in the New York Times, Boston Globe and Süddeutscher Zeitung.

6.7. Deciding trace equivalence (Objectives 1, 3)

Participants: Vincent Cheval, Hubert Comon-Lundh, Stéphanie Delaune, Rémy Chrétien.

Most existing results focus on trace properties like secrecy or authentication. There are however several security properties, which cannot be defined (or cannot be naturally defined) as trace properties and require the notion of indistinguishably. Typical examples are anonymity, privacy related properties or statements closer to security properties used in cryptography.

In the framework of the applied pi-calculus [44], as in similar languages based on equational logics, indistinguishability corresponds to a relation called trace equivalence. Roughly, two processes are trace equivalent when an observer cannot see any difference between the two processes. Static equivalence applies only to observations on finite sets of messages, and does not take into account the dynamic behavior of a process, whereas trace equivalence is more general and takes into account this aspect.

6.7.1. Static equivalence.

As explained above, static equivalence is a cornerstone to provide decision procedures for observational equivalence.

Stéphanie Delaune, in collaboration with Mathieur Baudet and Véronique Cortier, has designed a generic procedure for deducibility and static equivalence that takes as input any convergent rewrite system [15]. They have shown that their algorithm covers most of the existing decision procedures for convergent theories. They also provide an efficient implementation, and compare it briefly with the tools ProVerif and KiSs. This paper is a journal version of the work presented in [47].

In [17], Ştefan Ciobâcă, Stéphanie Delaune and Steve Kremer propose a representation of deducible terms to overcome the limitation of the procedure mentionned above. This new procedure terminates on a wide range of equational theories. In particular, they obtain a new decidability result for the theory of trapdoor bit commitment encountered when studying electronic voting protocols. The algorithm has been implemented in the KiSs tool. This paper is a journal version of the work presented in [49].

In [18], Stéphanie Delaune, in collaboration with Véronique Cortier (LORIA, France), shows that existing decidability results can be easily combined for any disjoint equational theories: if the deducibility and indistinguishability relations are decidable for two disjoint theories, they are also decidable for their union. They also propose a general setting for solving deducibility and indistinguishability for an important class (called *monoidal*) of equational theories involving AC operators. This paper is a journal version of the works presented in [45], [50].

6.7.2. Trace equivalence.

When processes under study do not contain replication, trace equivalence can be reduced to the problem of deciding symbolic equivalence, an equivalence relation introduced by M. Baudet [46].

Stéphanie Delaune, Steve Kremer, and Daniel Pasaila study this symbolic equivalence problem when cryptographic primitives are modeled using a group equational theory, a special case of monoidal equational theories. The results strongly rely on the correspondance between group theories and rings. This allows them to reduce the problem under study to the problem of solving systems of equations over rings. This result was published at IJCAR'12 [33],

When processes under study contain replication, the approach relying on symbolic equivalence does not work anymore. Moreover, since it is well-known that deciding reachability properties is undecidable under various restrictions, there is actually no hope to do better for equivalence-based properties. Rémy Chrétien, Véronique Cortier, and Stéphanie Delaune provide the first results of (un)decidability for certain classes of protocols for the equivalence problem. They consider a class of protocols shown to be decidable for reachability properties, and establish a first undecidability result. Then, they restrained the class of protocols a step further by making the protocols deterministic in some sense and preventing it from disclosing secret keys. This tighter class of protocols was then shown to be decidable after reduction to an equivalence between deterministic pushdown automata (see [42])

To deal with replication, another approach was studied by Vincent Cheval in collaboration with Bruno Blanchet. They propose an extension of the automatic protocol verifier ProVerif. ProVerif can prove observational equivalence between processes that have the same structure but differ by the messages they contain. In order to extend the class of equivalences that ProVerif handles, they extend the language of terms by defining more functions (destructors) by rewrite rules. These extensions have been implemented in ProVerif and allow one to automatically prove anonymity in the private authentication protocol by Abadi and Fournet. This work is currently under submission [40].

6.8. Mobile ad-hoc networks (Objectives 1, 3)

Participants: Rémy Chrétien, Stéphanie Delaune, Graham Steel.

Mobile ad hoc networks consist of mobile wireless devices which autonomously organize their communication infrastructure: each node provides the function of a router and relays packets on paths to other nodes. Finding these paths in an a priori unknown and constantly changing network topology is a crucial functionality of any ad hoc network. Specific protocols, called *routing protocols*, are designed to ensure this functionality known as *route discovery*. Secured versions of routing protocols have been proposed to provide more guarantees on the resulting routes, and some of them have been designed to protect the privacy of the users.

However, existing results and tools do not apply to routing protocols. This is due in particular to the fact that all possible topologies (infinitely many) have to be considered. Véronique Cortier, Jan Degrieck, and Stéphanie Delaune propose a simple reduction result: when looking for attacks on properties such as the validity of the route, it is sufficient to consider topologies with only four nodes, resulting in a number of just five distinct topologies to consider. As an application, several routing protocols, such as the SRP applied to DSR and the SDMSR protocols, have been analysed using the ProVerif tool. This work was published at POST'12 [32].

Rémy Chrétien and Stéphanie Delaune propose a framework for analysing privacy-type properties for routing protocols. They use the notion of equivalence between traces to formalise three security properties related to privacy, namely indistinguishability, unlinkability, and anonymity. They study the relationship between these definitions and we illustrate them using two versions of the ANODR routing protocol. This work is currently under submission [43].

In the context of vehicular ad-hoc networks, to improve road safety, a vehicle-to-vehicle communication platform is currently being developed by consortia of car manufacturers and legislators. In [51], Morten Dahl, Stéphanie Delaune and Graham Steel propose a framework for formal analysis of privacy in location based services such as anonymous electronic toll collection. They give a formal definition of privacy, and apply it to the VPriv scheme for vehicular services. They analyse the resulting model using the ProVerif tool, concluding that the privacy property holds only if certain conditions are met by the implementation. Their analysis includes some novel features such as the formal modelling of privacy for a protocol that relies on interactive zero-knowledge proofs of knowledge and list permutations.

6.9. Composition results (Objective 1)

Participants: Vincent Cheval, Stéphanie Delaune.

Formal methods have proved their usefulness for analysing the security of protocols. However, protocols are often analysed in isolation, and this is well-known to be not sufficient as soon as the protocols share some keys. Nowdays, several composition results exist for trace-based properties, but there is a lack of composition results for equivalence-based properties.

Myrto Arapinis, Vincent Cheval, and Stéphanie Delaune study the notion of trace equivalence and we show how to establish such an equivalence relation in a modular way. They show that composition works even when the processes share secrets provided that they satisfy some reasonable conditions. Their composition result allows one to prove various equivalence-based properties in a modular way, and works in a quite general setting. In particular, they consider arbitrary cryptographic primitives and processes that use non-trivial else branches. As an example, they consider the ICAO e-passport standard, and they show how the privacy guarantees of the whole application can be derived from the privacy guarantees of its sub-protocols. This work was published at CSF'12 [22].

TASC Project-Team

6. New Results

6.1. Constraint and Abstract Interpretation

Participants: Marie Pelleau, Charlotte Truchet, Fredéric Benhamou, Antoine Miné.

We apply techniques from Abstract Interpretation (AI), a general theory of semantic abstractions, to Constraint Programming (CP), which aims at solving hard combinatorial problems with a generic framework based on first-order logics. We highlight some links and differences between these fields: both compute fix-points by iteration but employ different extrapolation and refinement strategies; moreover, consistencies in Constraint Programming can be mapped to non-relational abstract domains.

- In a first step, we redefine all the components of CP on abstract domains, instead of the usual cartesian, domain-specific domains (boxes or integer sets), obtaining a generic method that can be specified for any of the AI abstract domains.
- In a second step, we then use the correspondences between AI and CP to build an abstract constraint solver that leverages abstract interpretation techniques (such as relational domains) to go beyond classic solvers. We present encouraging experimental results obtained with our prototype implementation, called AbSolute. In particular, AbSolute is able to solve problems on both discrete and continuous variables.

The work is done in collaboration with Antoine Miné.

A corresponding paper *A constraint solver based on abstract domains* [26] will appear at the 14th International Conference on Verification, Model Checking, and Abstract Interpretation (VMCAI'13).

6.2. Analytic Combinatorics and Lazy Filtering

Participants: Jérémie du Boisberranger, Danièle Gardy, Xavier Lorca, Charlotte Truchet.

The ANR Boole project (2009-2013) aims at quantifying different formats of boolean formulas, including SAT of constraints. Within the project, we have started a collaboration with Danièle Gardy, UVSQ, expert in analytic combinatorics and average-case study of algorithms. The goal of the collaboration was to quantify, within a high level probabilistic model, how often the bound-consistency propagator of an *alldifferent* constraint is likely to do something (or nothing). During year 2012, a particular focus has been put on calculating the probabilistic indicator, with an accepted publication at Analco 2013 (to appear). Further research include implementing and testing different possible uses for this indicator. A post-doc, Vincent Armant, has been recruited on the Boole project for this.

The corresponding paper *When is it worthwhile to propagate a constraint? A probabilistic analysis of* alldifferent [29] was accepted for publication at the ANALCO 13th Meeting on Analytic Algorithmics and Combinatorics (Analco 2013).

6.3. Learning Constraint Models

Participants: Nicolas Beldiceanu, Naina Razakarison, Helmut Simonis.

We designed a system which generates finite domain constraint models from positive example solutions, for highly structured problems. The system is based on the global constraint catalog, providing the library of constraints that can be used in modeling, and the constraint seeker tool, which finds a ranked list of matching constraints given one or more sample call patterns. We have tested the modeler with 230 examples, ranging from 4 to 6,500 variables, using between 1 and 7,000 samples. These examples come from a variety of domains, including puzzles, sports-scheduling, packing and placement, and design theory. Surprisingly, in many cases the system finds usable candidate lists even when working with *a single*, positive example.

The corresponding paper *A Model Seeker: Extracting Global Constraint Models From Positive Examples* [19] was published at the 18th International Conference on Principles and Practice of Constraint Programming (CP 2012).

6.4. Scalable Resource Scheduling Constraints

Participants: Nicolas Beldiceanu, Mats Carlsson, Arnaud Letort.

Following up on our work on scalable placement constraints for rectangle and box packing, and initially motivated by multidimensional bin packing problems that arise in the context of data centers, we have focussed this year our work on scalable resource scheduling constraints.

First we came up with a sweep based algorithm for the *cumulative* constraint, which can operate in filtering mode as well as in greedy assignment mode. Given n tasks, this algorithm has a worst-case time complexity of $O(n^2)$. In practice, we use a variant with better average-case complexity but worst-case complexity of $O(n^2 \log n)$, which goes down to $O(n \log n)$ when all tasks have unit duration, i.e. in the bin-packing case. Despite its worst-case time complexity, this algorithm scales well in practice, even when a significant number of tasks can be scheduled in parallel. It handles up to 1 million tasks in one single cumulative constraint in both CHOCO and SICStus.

Second we generalize the previous sweep algorithm to directly handle multiple resources. Given n tasks and k resources, this algorithm has a worst-case time complexity of $O(k \cdot n^2)$ but scales well in practice. In greedy assignment mode, it handles up to 1 million tasks with 64 resources in one single constraint. In filtering mode, on our benchmarks, it yields a speed-up of about $k^{0.75}$ when compared to its decomposition into k independent *cumulative* constraints.

A first paper *A Scalable Sweep Algorithm for the cumulative Constraint* [24] was published at the 18th International Conference on Principles and Practice of Constraint Programming (CP 2012). A second paper *A Synchronized Sweep Algorithm for the k-dimensional cumulative Constraint* was accepted for publication at the 10th International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems (CPAIOR 2013).

6.5. Reification of Global Constraints

Participants: Nicolas Beldiceanu, Mats Carlsson, Pierre Flener, Justin Pearson.

Being able expressing the negation of global constraints is something that is required in contexts such as testing the equivalence of two constraints models (see the PhD thesis of N. Lazaar) or in the context of learning constraints. Motivated by that, we introduce a simple idea for deriving reified global constraints in a systematic way. It is based on the observation that most global constraints can be reformulated as a conjunction of total function constraints together with a constraint that can be easily reified.

The corresponding paper *On the Reification of Global Constraints* [12] was published in the Constraints journal. A companion technical report [35] provides such reifications for 82% of the constraints of the global constraint catalog [36].

6.6. Optimization and Soft Problems

Participant: Thierry Petit.

Many optimization problems involve business constraints, which are complementary to an objective function that aggregates cost variables. These constraints involve the same cost variables. They are generally non linear. In the literature, several approaches were proposed for balancing constraints. We address the reverse concept, that is, concentrating high cost values in a restricted number of areas. This concept is motivated by several concrete examples, such as resource constrained scheduling problems with machine rentals. We present a new global constraint called *focus*. We provide a complete and optimum time complexity filtering algorithm for our constraint.

The corresponding paper *Focus : A Constraint for Concentrating High Costs* [27] was published at the 18th International Conference on Principles and Practice of Constraint Programming (CP 2012).

6.7. Consistency and Filtering

Participants: Nicolas Beldiceanu, Mats Carlsson, Gilles Chabert, Sophie Demassey, Thierry Petit, Jean-Charles Régin.

Following up on our work on efficient filtering algorithms for common conjunctions of widely used constraints (e.g., *among*, *alldifferent*, *linear constraint*, *inequalities constraints*) we provide:

1. An $O(n \log n)$ bound consistency filtering algorithm for the conjunction of an *alldifferent* and a *linear inequality* constraint. The $O(n \log n)$ complexity is equal to the complexity of the bound consistency algorithm of the *alldifferent* constraint.
2. A polynomial time bound consistency algorithm for the conjunction of *among* constraints where the variable and value domains are interval.

Motivated by the need to define more formally incomplete filtering algorithms we have proposed a new theoretical scheme for characterizing, comparing and classifying the intermediary levels of consistency of global constraints.

The corresponding papers, *An $O(n \log n)$ Bound Consistency Algorithm for the Conjunction of an alldifferent and an Inequality between a Sum of Variables and a Constant, and its Generalization* [17], *The Conjunction of Interval* among *Constraints* [21] and *Intermediary Local Consistencies* [28] were published at the 20th European Conference on Artificial Intelligence (ECAI 2012) as well as at the 9th International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems (CPAIOR 2012).

6.8. Automata and Matrix Models

Participants: Nicolas Beldiceanu, Mats Carlsson, Pierre Flener, Justin Pearson.

Matrix models are ubiquitous for constraint problems. Many such problems have a matrix of variables $\mathcal{M}$, with the same constraint C defined by a finite-state automaton $\mathcal{A}$ on each row of $\mathcal{M}$ and a global cardinality constraint gcc on each column of $\mathcal{M}$. We give two methods for deriving, by double counting, necessary conditions on the cardinality variables of the gcc constraints from the automaton $\mathcal{A}$. The first method yields linear necessary conditions and simple arithmetic constraints. The second method introduces the *cardinality automaton*, which abstracts the overall behaviour of all the row automata and can be encoded by a set of linear constraints. We also provide a arc-consistency filtering algorithm for the conjunction of lexicographic ordering constraints between adjacent rows of $\mathcal{M}$ and (possibly different) automaton constraints on the rows. We evaluate the impact of our methods in terms of runtime and search effort on a large set of nurse rostering problem instances.

The corresponding paper *On Matrices, Automata, and Double Counting in Constraint Programming* [11] was published in the Constraints journal.

6.9. Parallelization

Participants: Salvador Abreu, Yves Caniou, Philippe Codognet, Daniel Diaz, Florian Richoux.

During these last decades, many sequential algorithms for Constraint Satisfaction Problems (CSP) have been developed to be able to solve real problems from industry. However these problems become more and more complex and it remains important to treat them as fast as possible. Until the mid-2000's, one developed computers power by increasing CPU frequency. Nevertheless for about five years, this solution is not possible anymore since it asks too much energy (problem linked to heat dissipation issues), thus our machines architecture turns to be more and more multi-core oriented.

Nowadays we still have very few algorithms for constraint problems adapted to multi-core architecture. This year, we obtained very good results with the parallelization of meta-heuristic methods, reaching linear speed-ups over 8,192 cores on the *Costas Array Problem* [22], [23]. We also proposed in [20] two ways to perform smart cooperations between parallel local search processes, leading to very promising new approaches to solve constraint-based problems in parallel.

TOCCATA Team

6. New Results

6.1. Proofs of (Imperative) Programs

- A. Charguéraud has extended his ICFP'11 paper [70] into a journal paper, which is currently under review. This paper describes in more details the theory of characteristic formulae and the tool *CFML*, which supports the verification of *OCaml* programs through interactive *Coq* proofs.
- J.-C. Filliâtre has verified a two lines C program (solving the N-queens puzzle) using *Why3*. This case study has been presented at VSTTE 2012 [27].
- With M .Pereira and S. Melo de Sousa (Universidade da Beira Interior, Covilhã, Portugal), J.-C. Filliâtre developed an environment for proving ARM assembly code. It uses *Why3* as an intermediate VC generator. It was presented at the Inforum conference [34] (best student paper).
- F. Bobot and J.-C. Filliâtre have presented the notion of separation predicates introduced in the PhD of F. Bobot (defended December 2011) at ICFEM 2012 [21].
- S. Conchon and A. Mesbout, in collaboration with F. Zaïdi (Fortesse team, LRI) and A. Goel and S. Krstić (Strategic Cad Labs, INTEL), have presented a tool paper about the Cubicle model checker at CAV 2012 [24]. A more detailed description of the main algorithms implemented in Cubicle will be presented during the JFLA 2013 [73].
- A significant effort was dedicated to the development of *Why3*, with 3 public releases [39], [40], [41]. Associated with this activity, we actively participe to the new trend (that emerged in 2010-2011) of construction of international program verification benchmarks and organization of program verification competitions. We participated to the joint paper that reports on the first FoVeOOS competition [23] (http://proval.lri.fr/gallery/cost11comp.en.html). J.-C. Filliâtre and A. Paskevich wrote a detailed report [33] on the 2nd competition VSTTE competition (https://sites.google.com/site/vstte2012/compet) that they organized, published in the proceedings of the COMPARE workshop. This paper describes the competition, presents the five problems that were proposed to the participants, and gives an overview of the solutions sent by the 29 teams that entered the competition.

  Our own gallery of verified programs (http://toccata.lri.fr/gallery/index.en.html) was augmented significantly, with now approximately 100 examples, classified by topics, tools, etc.

6.2. Floating-Point and Numerical Programs

- The PhD thesis of T. Nguyen was defended in June [12]. It includes an improved version of the former approach [102] that we proposed for proving floating-point programs while taking into account architecture- and compiler-dependent features, such as the use of the x87 stack in Intel micro-processors. The underlying tool analyzes the assembly code generated by the compiler. It also includes a preliminary and independent approach for proving floating-point programs involving bit-level operations.
- C. Lelay, under the supervision of S. Boldo and G. Melquiond, has worked on easing proofs of differentiability and integrability in *Coq*. The use case was the existence of a solution to the wave equation thanks to D'Alembert's formula; the goal was to automate the process as much as possible [30]. While a major improvement with respect to *Coq* standard library, this first approach was not user-friendly enough for parametric intervals. So a different approach based on the pervasive use of total functions has been experimented with [22].
- S. Boldo, F. Clément, J.-C. Filliâtre, M. Mayero, G. Melquiond and P. Weis finished the formal proof of a numerical analysis program: the second order centered finite difference scheme for the one-dimensional acoustic wave [14].

- S. Boldo has developed a formal proof of an algorithm for computing the area of a triangle, an improvement of its error bound and new investigations in case of underflow [60].
- S. Boldo, J.-H. Jourdan, X. Leroy, and G. Melquiond have extended CompCert to get the first formally verified compiler that provably preserves the semantics of floating-point programs [63].
- G. Melquiond has kept improving the floating-point and interval theories used to perform proofs by computations in *Coq* [16].

6.3. Automated Deduction

- In collaboration with Assia Mahboubi (from Typical Inria project-team), and Guillaume Melquiond, the group involved in the development of *Alt-Ergo*, implemented and proved the correctness of a novel decision procedure for quantifier-free linear integer arithmetic [20]. This algorithm tries to bridge the gap between projection and branching/cutting methods: it interleaves an exhaustive search for a model with bounds inference. These bounds are computed provided an oracle capable of finding constant positive linear combinations of affine forms. An efficient oracle based on the Simplex procedure has been designed. Our algorithm is proved sound, complete, and terminating and is implemented in the *Alt-Ergo* theorem prover.
- In their LMCS journal paper [15], S. Conchon, É. Contejean and M. Iguernelala present a modular extension of ground AC-completion for deciding formulas in the combination of the theory of equality with user-defined AC symbols, uninterpreted symbols and an arbitrary signature disjoint Shostak theory X. This paper extends the results presented in [72] by showing that a simple preprocessing step allows to get rid of a full AC-compatible reduction ordering, and to simply use a partial multiset extension of a *non necessarily AC-compatible* ordering.
- In [31], S. Conchon, G. Melquiond and C. Roux described a dedicated procedure for a theory of floating-point numbers which allows reasoning on approximation errors. This procedure is based on the approach of the Gappa tool: it performs saturation of consequences of the axioms, in order to refine bounds on expressions. In addition to the original approach, bounds are further refined by a constraint solver for linear arithmetic. This procedure has been implemented in *Alt-Ergo*.
- In [42], [32], C. Dross and J. Kanig from AdaCore, in collaboration with S. Conchon and A. Paskevich propose a generic framework for adding a decision procedure for a theory or a combination of theories to an SMT prover. This mechanism is based on the notion of instantiation patterns, or *triggers*, which restrict instantiation of universal premises and can effectively prevent a combinatorial explosion. A user provides an axiomatization with triggers, along with a proof of completeness and termination in our framework, and obtains in return a sound, complete and terminating solver for his theory. A prototype implementation was realized in the Alt-Ergo prover. As a case study, a feature-rich axiomatization of doubly-linked lists was proved comlpete and terminating.
- In [38], A. Paskevich in collaboration with J. Blanchette from TU München, introduced a new format in the TPTP family (http://tptp.org), called TFF1, which extends the earlier TFF0 format (many-sorted first-order logic) with rank-1 type polymorphism. The technical report presents the syntax, typing rules, and semantics, as well as a sound and complete translation from TFF1 to TFF0. The format is designed to be easy to process by existing reasoning tools that support ML-style polymorphism. It opens the door to useful middleware, such as monomorphizers and other translation tools that encode polymorphism in FOF or TFF0. Ultimately, the hope is that TFF1 will be implemented in popular automatic theorem provers.
- A. Paskevich and J.-C. Filliâtre implemented a new *Coq* tactic that is able call an automated prover from *Coq* environment. It uses *Why3* as an intermediate tool. This new tactic brings a very significant improvement of proof automation within *Coq*. For example, the development of a certified VC generator in *Why3* made an intensive use of this tactic. The combination of automatic and interactive theorem proving was the subject of invited talks given by J.-C. Filliâtre at the workshop “Automation in Proof Assistants” [17] (satellite workshop of ETAPS 2012) and at the international workshop on Intermediate Verification Languages [18] (BOOGIE 2012, Berkeley, California, USA, July 2012).

- Together with O. Hermant (ISEP, Paris), D. Cousineau studied the cut elimination property for deduction modulo theories. They were able to show a strong relationship the syntactic cut-elimination property and the semantic construction of pre-models: they made a full semantic proof that the existence of a pre-model entails the cut elimination property for the considered theory in deduction modulo. This is published at the RTA Conference [26].
- *TLA+* is a specification language based on standard set theory and temporal logic, developed by the TLA groupe of Microsoft Research (http://research.microsoft.com/en-us/um/people/lamport/tla/tla.html). During the first part of his post-doc, D. Cousineau finalized a work on describing how to write *TLA+* proofs and check them with *TLAPS*, the *TLA+* Proof System. It was published as a tool description at FM Conference [25].
- S. Conchon defended his *habilitation à diriger des recherches* in December 2012. The memoir [11] provides a very good and useful survey of the scientific work of the past 10 years, around the SMT solving techniques, that leaded to the tools *Alt-Ergo* and Cubicle as they are nowadays.

6.4. Certification

- P. Herms, together with C. Marché and B. Monate (CEA List), developed a certified VC generator, using Coq. The program for VC calculus and its specifications are both written in Coq, but the code is crafted so that it can be extracted automatically into a stand-alone executable. It is also designed in a way that allows the use of arbitrary first-order theorem provers to discharge the generated obligations [28].
- On top of the previous generic VC generator, P. Herms developed a certified VC generator for C source code annotated using ACSL. This work is the main result of his PhD thesis which will be defended in January 2013.
- A. Tafat and C. Marché started experiments of development of a certified VC generator using Whyt instead of Coq. The challenge was to formalize the operational semantics of an imperative language, and a corresponding weakest precondition calculus, without the possibility to use Coq advanced features such as dependent types nor higher-order functions. The classical issues with local bindings, names and substitutions were solved by identifying appropriate lemmas. It was shown that *Why3* can offer a very significantly higher amount of proof automation compared to Coq [43]. This will be presented at the JFLA conference in February 2013 [95]
- The work that we started in 2011, about the use of the *Why3* environment and its back-end provers as an alternative to the built-in prover of "Atelier B", was published at the ABZ conference [29]. This work continues in the context of the new ANR project BWare.
- With J. Almeida, M. Barbosa, J. Pinto and B. Vieira (University do Minho, Braga, Portugal), J.-C. Filliâtre developed a method for certifying programs involving cryptographic methods. It uses *Why* as an intermediate language. A journal article will appear on *Science of Computer Programming* [13].
- Watermarking techniques are used to help identify copies of publicly released information. They consist in applying a slight and secret modification to the data before its release, in a way that should remain recognizable even in (reasonably) modified copies of the data. Using the *Coq*ALEA library, which formalizes probability theory and probabilistic programs, D. Baelde together with P. Courtieu, D. Gross-Amblard from Rennes and C. Paulin have established new results about the robustness of watermarking schemes against arbitrary attackers. The technique for proving robustness is adapted from methods commonly used for cryptographic protocols and our work illustrates the strengths and particularities of the induced style of reasoning about probabilistic programs. This work has been presented at the conference ITP 2012 [19].
- Supervised by J. Falcou and C. Paulin during his M2 internship, N. Lupinski developed a formalisation of a skeleton language for automated generation of parallel programs. A kernel of the language has been identified, its semantics has been formalised in *Coq* where a construction is interpreted by a

relation between lists of entries and lists of outputs. A transformation scheme from the skeleton language towards JOCaml programs has been proposed and proven correct with respect to the relational semantics. This work is described in [44].

- A. Charguéraud is currently working on the JsCert project (http://jscert.org), which aims at the formalization of the semantics of the JavaScript programming language (as described in *ECMAScript Language Specification, version 5.1*) and the development of a verified JavaScript interpreter. This project is joint work with Philippa Gardner, Sergio Maffeis, Gareth Smith, Daniele Filaretti and Daiva Naudziuniene from Imperial College, and Alan Schmitt and Martin Bodin from Inria Rennes - Bretagne Atlantique. As of today, the formalization already covers a substantial amount of the JavaScript language, and the verified interpreter is able to execute a number of benchmarks taken from standard JavaScript test suites.

  The formalization of the semantics of JavaScript makes use of a novel technique, called *pretty-big-step semantics*, for representing reduction rules in big-step style without suffering from a duplication of several premises accross different rules. This duplication is indeed typical in big-step semantics describing the behavior of exceptions and of divergence. The pretty-big-step semantics is described by A. Charguéraud in a paper to appear at ESOP 2013 [71].

TRIO Project-Team

6. New Results

6.1. Evaluation and optimal dimensioning of real-time systems

- **Code analyses and advanced visualization of software in real-time**

  Participants: Pierre Caserta, Olivier Zendra

  Last years, strong developments for our instrumentation, tracer and analyzer, had been performed, allowing us to really enter the experimental phase and getting first interesting results. A thorough state of the art had also been written.

  This state of the art paper had finally been published in TVCG, a leading journal in computer visualization. Thanks to the experimental setup efforts of previous years, we had been able in 2011 to conduct good experiments. We had designed and implemented a new way to visualize relations between software elements. These relations include static relations (is-a, direct heir, caller, callee, etc.) and dynamic ones (runtime caller, runtime callee). Our new relation visualization comprises a new way of placing way points so as to significantly decrease spatial and visual clutter when visualizing software systems with large numbers (thousands up to millions) of relations. This had lead to a publication in VISSOFT, one of the most recognized conferences in the software visualization domain, as well as a Best Poster in ECOOP, one of the most recognized conferences in the object-oriented domain. The important design and implementation work we had realized on the tracing and analysis software also lead to the publication of our method in ICOOOLPS 2011.

  This year, in 2012, we published our instrumentation and tracing method in Elsevier's Science of Computer Programming journal [9].

  Work has been going onto analyze polymorphism in Java programs, answering an apparently simple yet so far unanswered question: how much polymorphism is there actually in Java programs. This is of paramount importance, since a lot of work occur around polymorphism, which is an important concept, but no one is currently able to tell how much it impact programs in real life. We have begun writing this paper in cooperation with the LIRMM lab in Montpellier. In addition, we are in the process of finishing work pertaining to analyzing program evolutions, looking at differences between versions, and analyzing how dynamic metrics and static metrics correlate to evolution rate.

  Work in this domain has also lead to the writing and successful defense of Pierre Caserta's PhD thesis, titled "Analyse statique et dynamique de code et visualisation des logiciels via la métaphore de la ville : contribution à l'aide à la compréhension des programmes", on 7th December 2012 [7].

  A web site was also designed to publicize our work on the VITRAIL project.

- **Open Power and Energy Optimization PLatform and Estimator**

  Participants: Fabrice Vergnaud, Jérôme Vatrinet, Kévin Roussel, Olivier Zendra.

  Work in this domain was performed in the context of the ANR Open-PEOPLE (Open Power and Energy Optimization PLatform and Estimator) project, financed since the very end of 2008. Inria Nancy Grand Est is responsible for the software part of the platform and is involved in memory management for low-power issues. Work in this project begun in April 2009 (kick-off meeting). We have finished setting up the very important infrastructure for the software part of the Open-PEOPLE platform. We have finished expressing the requirements for the platform, in order to start the actual developments and the actual integration of tools provided by the different partners. In 2011, we have finished expressing the platform architecture and user interface (GUI). We have also finished implementing the part of the software platform that is the remote control to the hardware platform. We finally have finished implementing the core of the software platform and canonical models handling. This work lead to several technical and the several presentations and posters in conferences.

This year was the result harvesting for our project, in terms of development. We finished the design and implementation of the PCMD (Power Consumption Model Development) and the PCAO (Power Consumption Analysis and Optimization) parts of the software platform, as well as the external tools integration work. We also designed and implemented the Open-PEOPE model sharing website. Again, several demos and publications in conferences resulted [13], [21], [22].

- **Operator calculus and conception of algorithms for optimisatio of multi-constraints problems**

  Participants: Jamila Ben Slimane, Hugo Cruz-Sanchez, Bilel Nefzi, René Schott, Ye-Qiong Song

  R. Schott and G. Stacey Staples proposed a solution based on operator calculus for graphs with multi-constraints [26]. These constraints are not necessarily linear or positive. This approach was developed for realistic problems like:

  – configuration of satellites proposing a high-quality coverage [14];

  – optimal utilisation of ressources in hospitals;

  – optimal management in sensor networks [25].

    This work was the result of the collaboration of our team with MADYNES team, LPMA (Laboratoire de Probabilités et Modèles Aléatoires, Paris 6 et 7) and University of Illinois at Edwardsville.

6.2. Real-time analysis

- **Scheduling of tasks in automotive multicore ECUs**

  Participants: Aurélien Monot, Nicolas Navet, Françoise Simonot-Lion.

  As the demand for computing power is quickly increasing in the automotive domain, car manufacturers and tier-one suppliers are gradually introducing multicore ECUs in their electronic architectures. Additionally, these multicore ECUs offer new features such as higher levels of parallelism which ease the respect of safety requirements such as the ISO 26262 and the implementation of other automotive use-cases. These new features involve also more complexity in the design, development and verification of the software applications. Hence, car manufacturers and suppliers will require new tools and methodologies for deployment and validation. We address the problem of sequencing numerous elementary software components, called runnables, on a limited set of identical cores. We show how this problem can be addressed as two sub-problems, partitioning the set of runnables and building the sequencing of the runnables on each core, which problems cannot be solved optimally due to their algorithmic complexity. We then present low complexity heuristics to partition and build sequencer tasks that execute the runnable set on each core, and derive lower bounds on their efficiency (i.e., competitive ratio). Finally, we address the scheduling problem globally, at the ECU level, by discussing how to extend this approach in the case where other OS tasks are scheduled on the same cores as the sequencer tasks. An article providing a summary of this line of work has been published in IEEE TII [12].

- **Probabilistically analysable real-time system**

  Participants: Liliana Cucu-Grosjean, Adriana Gogonel, Codé Lo, Luca Santinelli, Dorin Maxim and Cristian Maxim.
  The adoption of more complex hardware to respond to the increasing demand for computing power in next- generation systems exacerbates some of the limitations of static timing analysis for the estimation of the worst-case execution time (WCET) estimation. In particular, the effort of acquiring (1) detail information on the hardware to develop an accurate model of its execution latency as well as (2) knowledge of the timing behaviour of the program in the presence of varying hardware conditions, such as those dependent on the history of previously executed instructions. These

problems are also known as the timing analysis walls. The probabilistic timing analysis, a novel approach to the analysis of the timing behaviour of next-generation real-time embedded systems, provides answers to timing analysis walls. In [10], [15] we have showed how the probabilistic timing analysis attacks the timing analysis walls. We have also presented experimental evidence that shows how probabilistic timing analysis reduces the extent of knowledge about the execution platform required to produce probabilistically-safe and tight WCET estimations.

Based on existing estimations of WCET or minimal inter-arrival time, we may propose different probabilistic schedulability analyses [19], [11].

- **Statistical analysis of real-time systems**

  Participants: Liliana Cucu-Grosjean, Adriana Gogonel, Lu Yue, Thomas Nolte [Malardelan University], Rob Davis, Ian Bate [University of York], Michael Houston, Guillem Bernat [Rapita].

  The response time analysis of real-time systems usually needs the knowledge of WCET estimation and this knowledge is not always available, e.g., because of intelectual property issues. This problem may be avoided by estimating statistically either the WCET of a task [18], the inter-arrival time [17] or the response time of each task [23].

- **Probabilistic Component-based ApproachesParticipants:** Luca Santinelli, Patrick Meumeu Yomsi, Dorin Maxim, Liliana Cucu-Grosjean.

  We have proposed a probabilistic component-based model which abstracts in the interfaces both the functional and non-functional requirements of such systems. This approach allows designers to unify in the same framework probabilistic scheduling techniques and compositional guarantees that go from soft to hard real- time. We have provided sufficient schedulability tests for task systems using such framework when the scheduler is either preemptive fixed-priority or earliest deadline first. These results were published in [16].

TYPICAL Project-Team

5. New Results

5.1. Feit-Thompson

The Feit-Thompson is an important theorem stating that every finite group of odd order is solvable. It is an important step in the classification of finite groups. Its proof is remarkable through its difficulty and its length (more than 1000 pages of dense mathematical text).

This proof was entirely formalized in Coq. This effort was started six years ago, as a joint project of the project teams Typical, Marelle (Sophia-Antipolis) and the Inria-MSR joint center, under the supervision of Georges Gonthier. The proof was finished in september 2012 and is considered a remarkable achievement. It also gave birth to several side products, such as enhancements of the SSReflect proof language. For Typical, Assia Mahboubi, Enrico Tassi and Cyril Cohen were instrumental in this effort.

5.2. Formal Semantics of Type Theory

Bruno Barras finished an extensive formalization of Coq's type theory in Coq, as well as a large formalization of set theory. This work includes several new results and insights in the study of Type Theory and is the body of Barras' habilitation thesis to be defended early in 2013.

5.3. Study of Type Theories

Bruno Barras finished an extensive formalization of Coq's type theory in Coq, as well as a large formalization of set theory. This work includes several new results and insights in the study of Type Theory and is the body of Barras' habilitation thesis to be defended early in 2013.

Chantal Keller, with Marc Lasson, has presented a notion of parametricity in impredicative type theories, which yields some possible application in proof search [18].

5.4. Formal and computable algebra

Cyril Cohen and Assia Mahboubi have worked on representing various algebraic objects in Coq, in a way that allows computation. In particular, Cohen proposed and developed a representation of algebraic numbers in Coq, as presented in [16]. Assia Mahboubi has collaborated with Frédéric Chyzak (Inria Paris-Roquencourt, Algo team) on the certification of algorithms for D-finite objects.

5.5. Certifiable real optimization

Under the joint supervision of Stéphane Gaubert and Benjamin Werner, with Xavier Allamigeon, Victor Magron is investigating ways to check difficult real inequalities, over bounded domains, in ways which can be re-checked by proof systems like Coq. One such algorithm, combining convex optimization and Max-plus techniques is submitted for publication at ECC 2013.

5.6. Binder representation in Coq

Benjamin Werner has developed a generic tree datatype in Coq, which can encode any language with fixed-arity operators with binders. The application towards smoother formal treatment of such languages is still in progress.

5.7. SMT and Coq

Chantal Keller has enhanced the performances of her SMT-Coq interface based automatic tactic. More precisely, the code has been made more modular which allowed:

- A first interfacing with the renowned Z3 SMT prover from Microsoft Research,
- Extending SMT-Coq to the theory of Coq's native 31 bits integers.

5.8. Automated decision procedures

Assia Mahboubi has woked with members of the Proval team on a new decision procedure for integer arithmetics now intergrated in the Alt-Ergo SMT solver. Assia Mahboubi has worked with Stéphane Lengrand and Mahfuza Farooque on the design of a sequent calculus with focussing and on the conception of a proof search strategy in this calculus which simulates the Davis-Putman-Logemann-Loveland algorithme modulo theory (DPPL(T)) which is implemented by modern SMT-solvers. An implementation developped by Stéphane Lengrand illustrate this approach on standard SMT benchmarks.

VEGAS Project-Team

5. New Results

5.1. Classical computational geometry

5.1.1. Complexity analysis of random geometric structures made simpler

Average-case analysis of data-structures or algorithms is commonly used in computational geometry when the more classical worst-case analysis is deemed overly pessimistic. Since these analyses are often intricate, the models of random geometric data that can be handled are often simplistic and far from "realistic inputs".

In a joint work with Olivier Devillers and Marc Glisse (Inria GEOMETRICA) [20], we presented a new simple scheme for the analysis of geometric structures. While this scheme only produces results up to a polylog factor, it is much simpler to apply than the classical techniques and therefore succeeds in analyzing new input distributions related to smoothed complexity analysis. We illustrated our method on two classical structures: convex hulls and Delaunay triangulations. Specifically, we gave short and elementary proofs of the classical results that n points uniformly distributed in a ball in R^d have a convex hull and a Delaunay triangulation of respective expected complexities $\widetilde{\Theta}(n^{((d+1)/(d-1))})$ and $\widetilde{\Theta}(n)$. We then prove that if we start with n points well-spread on a sphere, e.g. an (ϵ,κ)-sample of that sphere, and perturb that sample by moving each point randomly and uniformly within distance at most δ of its initial position, then the expected complexity of the convex hull of the resulting point set is $\widetilde{\Theta}(\sqrt(n)^{(1-1/d)}\delta^{-(d-1)/(4d)})$.

5.1.2. On the monotonicity of the expected number of facets of a random polytope

Let K be a compact convex body in R^d, let K_n be the convex hull of n points chosen uniformly and independently in K, and let $f_i(K_n)$ denote the number of i-dimensional faces of K_n.

In a joint work with Olivier Devillers and Marc Glisse (Inria GEOMETRICA) and Matthias Reitzner (Univ. Osnabruck) [21], we showed that for planar convex sets, $E(f_0(K_n))$ is increasing in n. In dimension $d \geq 3$ we prove that if $lim_{n\to\infty}\frac{E(f_{d-1}(K_n))}{An^c}=1$ for some constants A and $c>0$ then the function $E(f_{d-1}(K_n))$ is increasing for n large enough. In particular, the number of facets of the convex hull of n random points distributed uniformly and independently in a smooth compact convex body is asymptotically increasing. Our proof relies on a random sampling argument.

5.1.3. Embedding geometric structures

We continued working this year on the problem of embedding geometric objects on a grid of $\mathbb{R}^3$. Essentially all industrial applications take, as input, models defined with a fixed-precision floating-point arithmetic, typically doubles. As a consequence, geometric objects constructed using exact arithmetic must be embedded on a fixed-precision grid before they can be used as input in other software. More precisely, the problem is, given a geometric object, to find a similar object representable with fixed-precision floating-point arithmetic, where similar means topologically equivalent, close according to some distance function, etc. We are working on the problem of rounding polyhedral subdivisions on a grid of $\mathbb{R}^3$, where the only known method, due to Fortune in 1999, considers a grid whose refinement depends on the combinatorial complexity of the input, which does not solve the problem at hand. This project is joint work with Olivier Devillers (Inria Geometrica) and William Lenhart (Williams College, USA) who was in sabbatical in our team in 2012.

5.2. Non-linear computational geometry

5.2.1. Geometry of robotic mechanisms

Parallel manipulators are a family of mechanisms, the geometry of which is difficult to compute in general. The use of algebraic methods allowed us to describe precisely the geometry of the configurations of different specific parallel manipulators, in collaboration with researchers from the IRCCyN laboratory in Nantes.

More precisely, moving a parallel robot toward specific parametric values can break it. A challenge is to describe this set of singularities. This was adressed for a planar mechanism with three degrees of freedom in [16] and a spatial mechanism with six degrees of freedom in [12].

Then, a more challenging question arises naturally. Given a famılly of mechanisms parametrized by some construction variables, is it possible to find a mechanism that has no singularities? A method based on Gröbner bases was proposed in [17] for a specific family of planar parallel robot with two degrees of freedom.

5.2.2. Solving bivariate systems and topology of algebraic curves

In the context of our algorithm Isotop for computing the topology of algebraic curves [28], we study the bit complexity of solving a system of two bivariate polynomials of total degree d with integer coefficients of bitsize τ. We focus on the problem of computing a Rational Univariate Representation (RUR) of the solutions, that is, roughly speaking, a univariate polynomial and two rational functions which map the roots of the polynomial to the two coordinates of the solutions of the system.

We work on an algorithm for computing RURs with worst-case bit complexity in $O(d^8 + d^7\tau + d^5\tau^2)$ (where polylogarithmic factors are omitted). In addition, we show that certified approximations of the real solutions can be computed from this representation with $O(d^8 + d^7\tau)$ bit operations. It should be stressed that our algorithm is deterministic and that it makes no genericity assumption.

When $\tau \in O(d^2)$, this complexity decreases by a factor d^2 the best known upper bound for computing Rational Univariate Representations of such systems and it matches the recent best known complexity (Emeliyanenko and Sagraloff, 2012) for "only" computing certified approximations of the solutions. This shows, in particular, that computing RURs of bivariate systems is in a similar class of (known) complexity as computing certified approximations of *one* of the variables of its real solutions.

This work is on-going and is done in collaboration with Fabrice Rouillier (Inria Ouragan).

5.3. Combinatorics and combinatorial geometry

5.3.1. Multinerves and Helly numbers of acyclic families

The nerve of a family of sets is a simplicial complex that records the intersection pattern of its subfamilies. Nerves are widely used in computational geometry and topology, because the nerve theorem guarantees that the nerve of a family of geometric objects has the same topology as the union of the objects, if they form a good cover.

In a joint work with Éric Colin de Verdière (CNRS-ENS) and Grégory Ginot (Univ. Paris 6) we relaxed the good cover assumption to the case where each subfamily intersects in a disjoint union of possibly several homology cells, and we proved a generalization of the nerve theorem in this framework, using spectral sequences from algebraic topology. We then deduced a new topological Helly-type theorem that unifies previous results of Amenta, Kalai and Meshulam, and Matoušek. This Helly-type theorem is used to (re)prove, in a unified way, bounds on transversal Helly numbers in geometric transversal theory.

This work was presented at SoCG 2012 [18], where it received one of the two "best paper" awards.

5.3.2. Set systems and families of permutations with small traces

In a joint work with Otfried Cheong (KAIST, South Korea) and Cyril Nicaud (Univ. Marne-La-Vallée), we studied two problems of the following flavor: how large can a family of combinatorial objects defined on a finite set be if its number of distinct "projections" on any small subset is bounded? We first consider set systems, where the "projections" is the standard notion of trace, and for which we generalized Sauer's Lemma on the size of set systems with bounded VC-dimension. We then studied families of permutations, where the "projections" corresponds to the notion of containment used in the study of permutations with excluded patterns, and for which we delineated the main growth rates ensured by projection conditions. One of our motivations for considering these questions is the "geometric permutation problem" in geometric transversal theory, a question that has been open for two decades.

This work was published in the European Journal of Combinatorics [13].

5.3.3. Simplifying inclusion-exclusion formulas

Let $F = \{F_1, F_2, ..., F_n\}$ be a family of n sets on a ground set X, such as a family of balls in R^d. For every finite measure μ on X, such that the sets of F are measurable, the classical inclusion-exclusion formula asserts that $\mu(F_1 \cup F_2 \cup \bullet\bullet\bullet \cup F_n) = \sum_{I:\varnothing \neq I \subseteq [n]} (-1)^{|I|+1} \mu(\cap_{i\in I} F_i)$; that is, the measure of the union is expressed using measures of various intersections. The number of terms in this formula is exponential in n, and a significant amount of research, originating in applied areas, has been devoted to constructing simpler formulas for particular families F.

In a joint work with Jiří Matoušek, Pavel Paták, Zuzana Safernová and Martin Tancer (Charles Univ., Prague) [22] we provided the apparently first upper bound valid for an arbitrary F: we showed that every system F of n sets with m nonempty fields in the Venn diagram admits an inclusion-exclusion formula with $m^{O((log n)^2)}$ terms and with ± 1 coefficients, and that such a formula can be computed in $m^{O((log n)^2)}$ expected time. We also constructed systems of n sets on n points for which every valid inclusion-exclusion formula has the sum of absolute values of the coefficients at least $\Omega(n^{3/2})$.

VERIDIS Project-Team

6. New Results

6.1. Automated and Interactive Theorem Proving

6.1.1. Combination of decision procedures

Participants: Pascal Fontaine, Simon Halfon, Stephan Merz, Christoph Weidenbach.

SMT solvers, combination, decision procedures, theorem proving

We investigate the theoretical limits of combining decision procedures and reasoners, as these are important for the development of the veriT solver (see section 5.1). It has long been known that it is possible to extend any decidable language (subject to a minor requirement on cardinalities) with predicates described by a Bernays-Schönfinkel-Ramsey theory (BSR). A formula belongs to the BSR decidable fragment if it is a conjunction of universal, function-free formulas. As a consequence of this theoretical result, it is possible to extend a decidable quantifier-free language with sets and set operators, relations, orders and similar concepts. This can be used to significantly extend the expressivity of SMT solvers. In previous work, we generalized this result to the decidable first-order class of monadic predicate logic, and to the two-variable fragment. In subsequent joint work with Carlos Areces from Universidad Nacional de Córdoba, Argentina, we showed that two other important decidable fragments (namely the Ackermann fragment, and several guarded fragments) are also easily combinable. In 2012, we considered, in the same spirit, the combination of theories that are not necessarily decidable [18]. In particular, we considered combinations of decision procedures and refutationally complete semi-decision procedures, as well as black-box combinations of different refutationally complete theorem provers, together with finite model finders. These results in particular yield theoretical foundations for how FOL provers can be combined with SMT techniques in a black-box style of integration.

6.1.2. Using symmetries in SMT

Participants: Pascal Fontaine, Stephan Merz.

theorem proving, SMT solvers, decision procedures, symmetry

Methods exploiting problem symmetries have been very successful in several areas including constraint programming and SAT solving. We proposed similar techniques for enhancing the performance of SMT-solvers by detecting symmetries in the input formulas and using them to prune the search space of the SMT algorithm. These techniques are based on the concept of (syntactic) invariance by permutation of symbols. In 2011, we presented a technique restricted to constants but which exhibited impressive results for some categories of formulas [4]; this technique was quickly implemented in major SMT solvers, including CVC4 and Z3.

In 2012, we designed a more general approach, based on graph isomorphism, for symmetry detection in the SMT context. Experimental analysis indicates that many formulas from the SMT-LIB repository exhibit symmetries that are left unexploited by the previous techniques. Finding new techniques to exploit these is the subject of ongoing work with the University of Cordoba in Argentina; we expect that breaking those symmetries will yield again some significative efficiency improvement.

6.1.3. Encoding TLA+ proof obligations for SMT solvers

Participants: Stephan Merz, Hernán-Pablo Vanzetto.

system verification, SMT solving, TLA

The TLA$^+$ proof system TLAPS (see section 5.2) is being developed within a project at the MSR-Inria Joint Centre to which we contribute. Proof obligations that arise during the verification of typical TLA$^+$ specifications require reasoning about the principal TLA$^+$ data structures such as sets, functions, arithmetic, tuples, and records. None of the backend provers present in the initial versions of TLAPS was able to reason effectively about steps involving several of these features, and in 2011 we started developing an improved backend for translating TLA$^+$ proof obligations to SMT-Lib, the generic input language of SMT solvers. The main challenge was to design a sound translation from untyped TLA$^+$ to the multi-sorted first-order logic that underlies SMT-Lib, and our original proposal was based on deriving type assignments to TLA$^+$ expressions in a custom type system useful for SMT-Lib. This approach sometimes failed to derive types for subexpressions or required stronger typing assumptions than those required by the semantics of untyped TLA$^+$.

In 2012, based on a suggestion by Ken McMillan, we investigated a different approach whose main idea is to embed SMT sorts such as integers in the global universe of TLA$^+$ values, and to axiomatically define operations such as addition or multiplication on the image of that embedding. This approach effectively delegates type inference to the SMT solver and can therefore handle arbitrary TLA$^+$ expressions. However, it generates many quantified background axioms that may render SMT solvers ineffective, and we developed powerful pre-processing techniques for replacing quantified axioms by their required ground instances. The SMT backend in the current release of TLAPS is based on a hybrid approach to translation, where type inference is used whenever possible in order to obtain simpler SMT input. The two translation techniques have been published in 2012 [19], [20], and they have been validated over many case studies in TLAPS. For example, it enables proving the correctness of simple mutual-exclusion algorithms essentially without user interaction, and of the Paxos consensus algorithm in just 130 interactions, whereas a previous proof attempt using the traditional backend provers was unsuccessful.

6.1.4. Compression of SMT proofs

Participants: Pascal Fontaine, Stephan Merz.

theorem proving, SMT solvers, decision procedures, combination of decision procedures

Integrating an SMT solver in a certified environment such as TLAPS or an LF-style proof assistant requires the solver to output proofs. Unfortunately, those proofs may be quite large, and the overhead of rechecking the proof may account for a significant fraction of the proof time. In previous work, we proposed a technique for reducing the size of propositional proofs based on the analysis of resolution graphs, which were justified in an algebra of resolution. Unfortunately, the complexity of these techniques turned out to be prohibitive, but we proposed practical and efficient algorithms for more restricted compression techniques. We continue to develop this line of work with our partners at TU Wien.

6.1.5. Augmenting the Expressiveness of Spass

Participants: Evgeny Kruglov, Arnaud Fietzke, Daniel Wand, Christoph Weidenbach.

automated theorem proving, superposition, linear arithmetic, proof assistants

In 2012 we focused on bridging the gap between the input logic of SPASS and more expressive logics as they are used by systems supporting full-fledged verification such as Isabelle and TLAPS. Main contributions were a specific version of an order-sorted language that can be eventually translated in a many-sorted logic. The latter is implemented in Spass in a prototypic way and first experiments showed significant improvements on proof obligations out of Isabelle/HOL. Actually, the enhancements allowed Spass to become the most powerful automated theorem proving system supporting Isabelle [14]. We are currently working on a coupling with TLAPS (see section 5.2).

A second important branch is the integration of arithmetic into SPASS and the development of the respective hierarchic superposition calculus. In the past [31], [38] we experimented with a black box integration of LP solvers and Z3 to delegate arithmetic reasoning tasks. Now we started our own white box implementation for linear arithmetic and could achieve significant speed-ups. Our own reasoning procedure, dedicated to the specific form of the arithmetic proof obligations generated by SPASS is 50 to 200 times faster than any black box integration [29]. On the calculus side we could prove hierarchic superposition modulo linear arithmetic to be a decision procedure for the ground case, thus strictly generalizing the DPLL(LA) set up, and to be a decision procedure [39], [40] for timed automata reachability and extensions thereof [17].

6.1.6. Verification of linear hybrid automata

Participant: Uwe Waldmann.

automated theorem proving, superposition, linear arithmetic, proof assistants

We propose an improved symbolic algorithm for the verification of linear hybrid automata with large discrete state spaces. Large discrete state spaces arise naturally in industrial hybrid systems, due to the need to represent discrete inputs, counters, sanity checkbits, possibly multiple concurrent state machines, system-degradation modes, and finite switching variables. To prove safety properties of such systems, it is necessary to combine techniques for analyzing a complex dynamic behaviour with state space exploration methods that can deal with hundreds of discrete variables. In our approach, we represent both the discrete part and the continuous part of the hybrid state space symbolically using a variant of AIGs (And-Inverter-Graphs). Key components of our method are redundancy elimination (to maintain a compact symbolic representation by deleting superfluous linear constraints) and constraint minimization (exploiting the fact that states already reached in previous iterations of the model-checking algorithm can be interpreted as "don't cares" in later steps). A journal article describing the technique appeared in Science of Computer Programming [9].

6.2. Proved development of algorithms and systems

6.2.1. Incremental development of distributed algorithms

Participants: Dominique Méry, Manamiary Andriamiarina.

distributed algorithms, refinement, verification, distributed protocols

The development of distributed algorithms and, more generally, of distributed systems, is a complex, delicate, and challenging process. The approach based on refinement helps to gain formality by using a proof assistant, and proposes to apply a design methodology that starts from the most abstract model and leads, in an incremental way, to the most concrete model, for producing a distributed solution. Our works help to formalize pre-existing algorithms, develop new algorithms, as well as develop models for distributed systems.

Our research was initially (until 2010) carried out within the ANR project RIMEL, in joint work with Mohammed Mosbah and Mohammed Tounsi from the LABRI laboratory, and we are maintaining a joint project B2VISIDIA with LABRI on these topics. More concretely, we aim at an integration of the correct-by-construction refinement-based approach into the *local computation* programming model. The team of LABRI develops an environment called VISIDIA that provides a toolset for developing distributed algorithms expressed as a set of rewriting rules of graph structures. The simulation of rewriting rules is based on synchronization algorithms and we have developed these algorithms by refinement.

More precisely, we show how state-based models can be developed for specific problems and how they can be simply reused by controlling the composition of state-based models through the refinement relationship. Consequently, we obtain a redevelopment of existing distributed algorithms in the *correct-by-construction* approach, and a framework for deriving new distributed algorithms (by integrating models) whose correctness is ensured by construction. Traditionally, distributed algorithms are supposed to run on a fixed network, whereas we consider a network with a changing topology. We have illustrated our methodology with the study of the protocol ANYCAST RP.

The contribution is related to the development of proof-based patterns providing effective help to the developer of formal models of applications, such as dynamic routing or the snapshot problem [13]. In fact, we have developed patterns for simplifying the development of distributed systems using refinement. The applicability of a pattern for routing has been reapplied to the development of a network on chip [12] with our partners of the French-Algerian cooperation described in section 8.3 .

6.2.2. Modeling and verifying the Pastry routing protocol

Participants: Tianxiang Lu, Stephan Merz, Christoph Weidenbach.

distributed hash table, peer-to-peer protocol, Pastry, model checking, theorem proving

As a significant case study for the techniques that we are developing within VeriDis, we are modeling and verifying the routing protocol of the Pastry algorithm [36] for maintaining a distributed hash table in a peer-to-peer network. As part of his PhD work, Tianxiang Lu has developed a TLA$^+$ model of the Pastry routing protocol, which has uncovered several issues in the existing presentations of the protocol in the literature, and in particular a loophole in the join protocol that had been fixed by the algorithm designers in a technical report that appeared after the publication of the original protocol.

As a first step towards proving correctness of the Pastry routing protocol, we identified in 2011 a number of candidate invariants and formally proved in TLAPS (see section 5.2) that these implied the high-level correctness property. In 2012, we consolidated these invariants and proved them correct for our model under the strong assumption that no node ever leaves the network, and the minor assumption that any active node can at any time only allow one new node to join the network. It is still not clear at the moment to which extent nodes can be allowed to leave the network without breaking the virtual ring maintained by Pastry. The invariant proofs contain almost 15000 interactions and constitutes the largest case study carried out so far using TLAPS. We have more recently been able to obtain better automation using the new SMT backend (see section 6.1). The proof was presented at the TLA workshop of FM 2012 [23].

6.2.3. Verification of distributed algorithms in the Heard-Of model

Participants: Henri Debrat, Stephan Merz.

theorem proving, distributed algorithms, round-based computation, Byzantine failures

Distributed algorithms are often quite subtle, both in the way they operate and in the assumptions required for their correctness. Formal models are important for unambiguously understanding the hypotheses and the properties of a distributed algorithm. We focus on the verification of round-based algorithms for fault-tolerant distributed systems expressed in the Heard-Of model of Charron-Bost and Schiper [37], and have previously established a reduction theorem that allows to pretend that nodes operate synchronously.

In 2012, we have consolidated our formal proofs in Isabelle/HOL. In particular, we have finished the formal proof of the reduction theorem within Isabelle, produced a generic encoding of the Heard-Of model as a locale in Isabelle/HOL, and used this representation for verifying six different Consensus algorithms: three algorithms tolerating benign failures and three others designed for malicious failures, such as corrupted values. Our Isabelle theories have been published at the Archive of Formal Proofs [27]. The proof of the reduction theorem required formalizing the notion of stuttering invariance, which can be of independent interest and that has also been accepted at the Archive of Formal Proofs [28].

As a significant extension of this work, we have studied the formal verification of probabilistic Consensus algorithms in the Heard-Of model, in particular the Ben-Or algorithm.

6.2.4. Model checking within SimGrid

Participants: Marie Duflot-Kremer, Stephan Merz.

model checking, distributed algorithms, message passing, communication primitives, partial-order reduction

For several years we have cooperated with Martin Quinson from the AlGorille project team on adding model checking capabilities to the simulation platform SimGrid for message-passing distributed C programs. The expected benefit of such an integration is that programmers can complement simulation runs by exhaustive state space exploration in order to detect errors such as race conditions that would be hard to reproduce by testing. As part of the thesis work of Cristián Rosa (defended in 2011), a stateless model checker was implemented within the SimGrid platform that can be used to verify safety properties of distributed C programs that communicate by message passing. The ongoing thesis of Marion Guthmuller builds upon this work and aims to extend it for verifying certain liveness properties. This requires rethinking the stateless design, as well as adapting the dynamic partial-order reduction algorithm that is essential to limiting the part of the state space that must actually be explored.

6.2.5. Modeling Medical Devices

Participant: Dominique Méry.

Formal modelling techniques and tools have attained sufficient maturity for formalizing highly critical systems in view of improving their quality and reliability, and the development of such methods has attracted the interest of industrial partners and academic research institutions. Building high quality and zero-defect medical software-based devices is a particular domain where formal modelling techniques can be applied effectively. In [21], we present a methodology for developing critical systems from requirement analysis to automatic code generation based on a standard safety assessment approach. This methodology combines refinement, proof, model checking, and animation, and ultimately can automatically generate source code. This approach is intended to contribute to further the use of formal techniques for developing critical systems with high integrity and to verify complex properties. An assessment of the proposed methodology is given through developing a standard case study: the cardiac pacemaker.

Medical devices are very prone to showing unexpected system behaviour in operation when traditional methods are used for system testing. Device-related problems have been responsible for a large number of serious injuries. Officials of the US Food and Drug Administration (FDA) found that many deaths and injuries related to these devices are caused by flaws in product design and engineering. Cardiac pacemakers and implantable cardioverter-defibrillators (ICDs) are among the most critical medical devices and require closed-loop modelling (integrated system and environment modelling) for verification purposes before obtaining a certificate from the certification bodies. In [24] we present a methodology for modelling a biological system, such as the heart. The heart model is based mainly on electrocardiography analysis, which provides a model at the cellular level. Combining this environment model with a formal model of the pacemaker, we obtain a closed-loop model over which the overall correctness can be verified.

Clinical guidelines systematically assist practitioners in providing appropriate health care in specific clinical circumstances. Today, a significant number of guidelines and protocols are lacking in quality. Indeed, ambiguity and incompleteness are likely anomalies in medical practice. In [25] we use the Event-B modeling language to represent guidelines for subsequent validation. Our main contributions are: to apply mathematical formal techniques to evaluate real-life medical protocols for quality improvement, to derive verification proofs for the protocol and properties according to medical experts, and to publicize the potential of this approach. An assessment of the proposed approach is given through a case study, relative to a real-life reference protocol concerning ECG interpretation, for which we uncovered several anomalies.

Finally, we propose a refinement-based methodology [10] for complex medical systems design, which possesses the required key features. A refinement-based combined approach of formal verification, model validation using a model-checker and refinement chart is proposed in this methodology for designing a high-confidence medical device. Furthermore, we show the effectiveness of this methodology for the design of a cardiac pacemaker system.

6.2.6. Fundamentals of Network Calculus in Isabelle/HOL

Participant: Stephan Merz.

networked systems, min-plus algebra, formal proof

The design of networked and embedded systems has traditionally been accompanied by formal methods for design and analysis. Network Calculus [42] is a well-established theory, based on the $(\min, +)$ dioid, that is designed for computing delay and memory bounds in networks. The theory is supported by several commercial and open-source tools and has been used in major industrial applications, such as the design and certification of the Airbus A380 AFDX backbone. Nevertheless, it is difficult for certification authorities to assess the correctness of the computations carried out by the tools supporting Network Calculus, and we propose the use of *result certification* techniques for increasing the confidence in the Network Calculus toolchain. In joint work with Marc Boyer from ONERA in Toulouse, and with Loïc Fejoz and Nicolas Navet from the RealTime at Work (RTaW) company, we have supervised the master thesis of Etienne Mabille to evaluate the feasibility of the approach. Parts of the theory underlying Network Calculus were formalized in the proof assistant Isabelle/HOL, and this encoding was used to formally derive theorems that underly the computation of bounds in network servers. The Network Calculus tool produced by RTaW was instrumented to generate traces of its computation, and the correctness of simple systems could in this way be certified by Isabelle. A publication of this work is in preparation, and we intend to continue and extend it in a future joint project.

6.2.7. Bounding message length in attacks against security protocols

Participant: Marie Duflot-Kremer.

security protocols, verification

Security protocols are short programs that describe communication between two or more parties in order to achieve security goals. Despite the apparent simplicity of such protocols, their verification is a difficult problem and has been shown to be undecidable in general. This undecidability comes from the fact that the set of executions to be considered is of infinite depth (an infinite number of protocol sessions can be run) and infinitely branching (the intruder can generate an unbounded number of distinct messages). Several attempts have been made to tackle each of these sources of undecidability. Together with Myrto Arapinis, we have shown [32] that, under a syntactic and reasonable condition of "well-formedness" on the protocol, we can get rid of the infinitely branching part. Following this conference publication, we are preparing a journal version of this result extending the set of security properties to which the result is applicable, in particular including authentication properties.

6.2.8. Evaluating and verifying probabilistic systems

Participant: Marie Duflot-Kremer.

verification, probabilistic systems, performance evaluation

Since its introduction in the 1980s, model checking has become a prominent technique for the verification of complex systems. The aim was to decide whether or not a system was fulfilling its specification. With the rise of probabilistic systems, new techniques have been designed to verify this new type of systems, and appropriate logics have been proposed to describe more subtle properties to be verified. However, some characteristics of such systems cannot fall in the field of model checking. The aim is thus not to tell wether a property is satisfied but how well the system performs with respect to a certain measure. Together with researchers from ENS de Cachan and University Paris Est Créteil we have designed a statistical tool made to tackle both performance and verification issues. Following several conference talks, a journal paper is currently written to present both the approach as well as application to a concrete case study: flexible manufacturing systems.

VERTECS Project-Team

6. New Results

6.1. Verification

6.1.1. Probabilistic ω-automata

Participant: Nathalie Bertrand.

Probabilistic ω-automata are a variant version of nondeterministic automata over infinite words where all choices are resolved by probabilistic distributions. Acceptance of a run for an infinite input word can be defined using traditional acceptance criteria for ω-automata, such as Büchi, Rabin or Streett conditions. The accepted language of a probabilistic ω-automata is then defined by imposing a constraint on the probability measure of the accepting runs. Together with Christel Baier and Marcus Grösser from TU Dresden, we studied a series of fundamental properties of probabilistic ω-automata with three different language-semantics: (1) the probable semantics that requires positive acceptance probability, (2) the almost-sure semantics that requires acceptance with probability 1, and (3) the threshold semantics that relies on an additional parameter λ in]0,1[that specifies a lower probability bound for the acceptance probability. We provided a comparison of probabilistic ω-automata under these three semantics and nondeterministic ω-automata concerning expressiveness and efficiency. Furthermore, we addressed closure properties under the Boolean operators union, intersection and complementation and algorithmic aspects, such as checking emptiness or language containment. This work was published in Journal of the ACM [6].

6.1.2. Petri nets reachability graphs

Participant: Christophe Morvan.

In the article [10], we investigate the decidability and complexity status of model-checking problems on unlabelled reachability graphs of Petri nets by considering first-order and modal languages without labels on transitions or atomic propositions on markings. We consider several parameters to separate decidable problems from undecidable ones. Not only are we able to provide precise borders and a systematic analysis, but we also demonstrate the robustness of our proof techniques.

6.1.3. Frequencies in timed automata

Participant: Amélie Stainer.

A quantitative semantics for infinite timed words in timed automata based on the frequency of a run was introduced earlier by Bertrand, Bouyer, Brihaye and Stainer. Unfortunately, most of the results are obtained only for one-clock timed automata because the techniques do not allow to deal with some phenomenon of convergence between clocks. On the other hand, the notion of forgetful cycle was introduced by Basset and Asarin, in the context of entropy of timed languages, and seems to detect exactly these convergences. In [20], we investigate how the notion of forgetfulness can help to extend the computation of the set of frequencies to n-clock timed automata.

6.1.4. Bounded satisfiability for PCTL

Participant: Nathalie Bertrand.

While model checking PCTL for Markov chains is decidable in polynomial-time, the decidability of PCTL satisfiability, as well as its finite model property, are long standing open problems. While general satisfiability is an intriguing challenge from a purely theoretical point of view, we argue that general solutions would not be of interest to practitioners: such solutions could be too big to be implementable or even infinite. Inspired by bounded synthesis techniques, we turn to the more applied problem of seeking models of a bounded size: we restrict our search to implementable – and therefore reasonably simple – models. In [14] and together with John Fearnley and Sven Schewe from University of Liverpool, we propose a procedure to decide whether or not a given PCTL formula has an implementable model by reducing it to an SMT problem. We have implemented our techniques and found that they can be applied to the practical problem of sanity checking – a procedure that allows a system designer to check whether their formula has an unexpectedly small model.

6.1.5. Graph transformation systems

Participant: Nathalie Bertrand.

In [13], we study decidability issues for reachability problems in graph transformation systems, a powerful infinite-state model. For a fixed initial configuration, we consider reachability of an entirely specified configuration and of a configuration that satisfies a given pattern (coverability). The former is a fundamental problem for any computational model, the latter is strictly related to verification of safety properties in which the pattern specifies an infinite set of bad configurations. In this paper we reformulate results obtained, e.g., for context-free graph grammars and concurrency models, such as Petri nets, in the more general setting of graph transformation systems and study new results for classes of models obtained by adding constraints on the form of reduction rules.

6.2. Active and passive testing

6.2.1. More testable properties

Participants: Thierry Jéron, Hervé Marchand.

Testing remains a widely used validation technique for software systems. However, recent needs in software development (e.g., in terms of security concerns) may require to extend this technique to address a larger set of properties. In [11], we explore the set of testable properties within the Safety-Progress classification where testability means to establish by testing that a relation, between the tested system and the property under scrutiny, holds. We characterize testable properties w.r.t. several relations of interest. For each relation, we give a sufficient condition for a property to be testable. Then, we study and delineate a fine-grain characterization of testable properties: for each Safety-Progress class, we identify the subset of testable properties and their corresponding test oracle. Furthermore, we address automatic test generation for the proposed framework by providing a general synthesis technique that allows to obtain canonical testers for the testable properties in the Safety-Progress classification. Moreover, we show how the usual notion of quiescence can be taken into account in our general framework, and, how quiescence improves the testability results. Then, we list some existing testing approaches that could benefit from this work by addressing a wider set of properties. Finally, we propose Java-PT, a prototype Java toolbox that implements the results introduced in this article.

6.2.2. Runtime enforcement of timed properties

Participants: Thierry Jéron, Hervé Marchand, Srinivas Pinisetty.

Runtime enforcement is a powerful technique to ensure that a running system respects some desired properties. Using an enforcement monitor, an (untrusted) input execution (in the form of a sequence of events) is modified into an output sequence that complies to a property. Runtime enforcement has been extensively studied over the last decade in the context of untimed properties. The paper [19], introduces runtime enforcement of timed properties. We revisit the foundations of runtime enforcement when time between events matters. We show how runtime enforcers can be synthesized for any safety or co-safety timed property. Proposed runtime enforcers are time retardant: to produce an output sequence, additional delays are introduced between the events of the input sequence to correct it. Runtime enforcers have been prototyped and our simulation experiments validate their effectiveness.

6.2.3. Test generation for tiles systems

Participants: Sébastien Chédor, Thierry Jéron, Christophe Morvan.

In [17] we explore test generation for Recursive Tile Systems (RTS) in the framework of the classical ioco testing theory. The RTS model allows the description of reactive systems with recursion, and is very similar to other models like Pushdown Automata, Hyperedge Replacement Grammars or Recursive State Machines. We first present an off-line test generation algorithm for Weighted RTS, a determinizable sub-class of RTS, and second, an on-line test generation algorithm for the full RTS model. Both algorithms use test purposes to guide test selection through targeted behaviours.

6.2.4. Partially observed recursive tiles systems

Participants: Sébastien Chédor, Hervé Marchand, Christophe Morvan.

The analysis of discrete event systems under partial observation is an important topic, with major applications such as the detection of information flow and the diagnosis of faulty behaviors. In [18] we consider recursive tile systems, which are infinite systems generated by a finite collection of finite tiles, a simplified variant of deterministic graph grammars. Recursive tile systems are expressive enough to capture classical models of recursive systems, such as the pushdown systems and the recursive state machines. They are infinite-state in general and therefore standard powerset constructions for monitoring do not always apply. We exhibit computable conditions on recursive tile systems and present non-trivial constructions that yield effective computation of the monitors. We apply these results to the classic problems of opacity and diagnosability.

6.2.5. Off-line test selection with test purposes for non-deterministic timed automata

Participants: Nathalie Bertrand, Thierry Jéron, Amélie Stainer.

The LMCS article [7] proposes novel off-line test generation techniques from non-deterministic timed automata with inputs and outputs (TAIOs) in the formal framework of the tioco conformance theory. In this context, a first problem is the determinization of TAIOs, which is necessary to foresee next enabled actions after an observable trace, but is in general impossible because not all timed automata are determinizable. This problem is solved thanks to an approximate determinization using a game approach. The algorithm performs an io-abstraction which preserves the tioco conformance relation and thus guarantees the soundness of generated test cases. A second problem is the selection of test cases from a TAIO specification. The selection here relies on a precise description of timed behaviors to be tested which is carried out by expressive test purposes modeled by a generalization of TAIOs. Finally, an algorithm is described which generates test cases in the form of TAIOs equipped with verdicts, using a symbolic co-reachability analysis guided by the test purpose. Properties of test cases are then analyzed with respect to the precision of the approximate determinization: when determinization is exact, which is the case on known determinizable classes, in addition to soundness, properties characterizing the adequacy of test cases verdicts are also guaranteed.

6.2.6. Monitor-based statistical model checking of timed systems

Participant: Amélie Stainer.

In [16], we present a novel approach and implementation for analysing weighted timed automata (WTA) with respect to the weighted metric temporal logic ($\mathrm{WMTL}_{\leq}$). Based on a stochastic semantics of WTAs, we apply statistical model checking (SMC) to estimate and test probabilities of satisfaction with desired levels of confidence. Our approach consists in the generation of deterministic monitors for formulas in $\mathrm{WMTL}_{\leq}$, allowing for efficient SMC by run-time evaluation of a given formula. By necessity, the deterministic observers are in general approximate (over- or under-approximations), but are most often exact and experimentally tight. The technique is implemented in the new tool CASAAL. that we seamlessly connect to Uppaal-smc. in a tool chain. We demonstrate the applicability of our technique and the efficiency of our implementation through a number of case-studies.

6.3. Control synthesis

6.3.1. Synthesis of opaque systems

Participant: Hervé Marchand.

Opacity is a security property formalizing the absence of (secret) information leakage. We address the problem of synthesizing opaque systems. A secret predicate S over the runs of a system G is opaque to an external user having partial observability over G, if he can never infer from the observation of a run of G that the run belongs to S. We choose to control the observability of events by adding a device, called a mask, between the system G and the users. We first investigate the case of static partial observability where the set of events the user can observe is fixed once and for all by a static mask. In this context, we show that checking whether a system is opaque is PSPACE-complete, which implies that computing an optimal static mask ensuring opacity is also a PSPACE-complete problem. Next, we introduce dynamic partial observability where the set of events the user can observe changes over time and is determined by a dynamic mask. We show how to check that a system is opaque w.r.t. to a dynamic mask and also address the corresponding synthesis problem: given a system G and secret states S, compute the set of dynamic masks under which S is opaque. Our main result is that the set of such masks can be finitely represented and can be computed in EXPTIME and that this is a lower bound. We also address the problem of computing an optimal mask. This work was published in FMSD [9].

6.3.2. Symbolic Supervisory Control of Infinite Transition Systems under Partial Observation using Abstract Interpretation

Participant: Hervé Marchand.

In the DEDS article [12], we propose algorithms for the synthesis of state-feedback controllers with partial observation of infinite state discrete event systems modelled by Symbolic Transition Systems. We provide models of safe memoryless controllers both for potentially deadlocking and for deadlock free controlled systems. The termination of the algorithms solving these problems is ensured using abstract interpretation techniques which provide an overapproximation of the transitions to disable. We then extend our algorithms to controllers with memory and to online controllers. We also propose improvements in the synthesis of controllers in the finite case which, to our knowledge, provide more permissive solutions than previously proposed in the literature. Our tool SMACS gives an empirical validation of our methods by showing their feasibility, usability and efficiency.

6.3.3. Playing optimally on timed automata with random delays

Participant: Nathalie Bertrand.

In [15], we marry continuous time Markov decision processes (CTMDPs) with stochastic timed automata into a model with joint expressive power. This extension is very natural, as the two original models already share exponentially distributed sojourn times in locations. It enriches CTMDPs with timing constraints, or symmetrically, stochastic timed automata with one conscious player. Our model maintains the existence of optimal control known for CTMDPs. This also holds for a richer model with two players, which extends continuous time Markov games. But we have to sacrifice the existence of simple schedulers: polyhedral regions are insufficient to obtain optimal control even in the single-player case.

ALEA Project-Team

6. New Results

6.1. Bayesian Nonparametric models for ranked data and bipartite graphs.

In [20], the author develops a novel Bayesian nonparametric model for random bipartite graphs. The model is based on the theory of completely random measures and is able to handle a potentially infinite number of nodes. It is shown that the model has appealing properties and in particular it may exhibit a power-law behavior. Posterior characterization, a generative process for network growth, and a simple Gibbs sampler for posterior simulation are derived. The model is shown to be well fitted to several real-world social networks.

In [21], we develop a Bayesian nonparametric extension of the popular Plackett-Luce choice model that can handle an infinite number of choice items. Our framework is based on the theory of random atomic measures, with the prior specified by a gamma process. We derive a posterior characterization and a simple and effective Gibbs sampler for posterior simulation. We develop a time-varying extension of our model, and apply it to the New York Times lists of weekly bestselling books.

6.2. A new model for polychotomous data

Multinomial logistic regression is one of the most popular models for modelling the effect of explanatory variables on a subject choice between a set of specified options. This model has found numerous applications in machine learning, psychology or economy. Bayesian inference in this model is non trivial and requires, either to resort to a Metropolis-Hastings algorithm, or rejection sampling within a Gibbs sampler. In [19], we propose an alternative model to multinomial logistic regression. The model builds on the Plackett-Luce model, a popular model for multiple comparisons. We show that the introduction of a suitable set of auxiliary variables leads to an Expectation-Maximization algorithm to find Maximum A Posteriori estimates of the parameters. We further provide a full Bayesian treatment by deriving a Gibbs sampler, which only requires to sample from highly standard distributions. We also propose a variational approximate inference scheme. All are very simple to implement. One property of our Plackett-Luce regression model is that it learns a sparse set of feature weights. We compare our method to sparse Bayesian multinomial logistic regression and show that it is competitive, especially in presence of polychotomous data.

6.3. Sparsity-Promoting Bayesian Dynamic Linear Models

Sparsity-promoting priors have become increasingly popular over recent years due to an increased number of regression and classification applications involving a large number of predictors. In time series applications where observations are collected over time, it is often unrealistic to assume that the underlying sparsity pattern is fixed. We propose in [37] an original class of flexible Bayesian linear models for dynamic sparsity modelling. The proposed class of models expands upon the existing Bayesian literature on sparse regression using generalized multivariate hyperbolic distributions. The properties of the models are explored through both analytic results and simulation studies. We demonstrate the model on a financial application where it is shown that it accurately represents the patterns seen in the analysis of stock and derivative data, and is able to detect major events by filtering an artificial portfolio of assets.

6.4. Evolutionnary algorithms and genetic programming

In [22], we consider the identification of a nonlinear system modelled by a nonlinear output error (NOE) model when the system output is disturbed by an additive zero-mean white Gaussian noise. In that case, standard on-line or off-line least squares methods may lead to poor results. Here, our approach is based on evolutionary algorithms. Although their computational cost can be higher than the above methods, these algorithms present some advantages, which often lead to an effortless optimisation. Indeed, they do not need an elaborate formalisation of the problem. When their parameters are correctly tuned, they avoid to get stuck at a local optimum. To take into account the influence of the additive noise, we investigate different approaches and we suggest a whole protocol including the selection of a fitness function and a stop rule. Without loss of generality, simulations are provided for two nonlinear systems and various signal-to-noise ratios.

The regularity of a signal can be numerically expressed using Hölder exponents, which characterize the singular structures a signal contains. In particular, within the domains of image processing and image understanding, regularity-based analysis can be used to describe local image shape and appearance. However, estimating the Hölder exponent is not a trivial task, and current methods tend to be computationally slow and complex. The paper [17] presents an approach to automatically synthesize estimators of the pointwise Hölder exponent for digital images. This task is formulated as an optimization problem and Genetic Programming (GP) is used to search for operators that can approximate a traditional estimator, the oscillations method. Experimental results show that GP can generate estimators that achieve a low error and a high correlation with the ground truth estimation. Furthermore, most of the GP estimators are faster than traditional approaches, in some cases their runtime is orders of magnitude smaller. This result allowed us to implement a real-time estimation of the Hölder exponent on a live video signal, the first such implementation in current literature. Moreover, the evolved estimators are used to generate local descriptors of salient image regions, a task for which a stable and robust matching is achieved, comparable with state-of-the-art methods. In conclusion, the evolved estimators produced by GP could help expand the application domain of Hölder regularity within the fields of image analysis and signal processing.

One of the main open problems within Genetic Programming (GP) is to meaningfully characterize the difficulty (or hardness) of a problem. The general goal is to develop predictive tools that can allow us to identify how difficult a problem is for a GP system to solve. In [23] and [24], we identify and compare two main approaches that address this question. We denote the first group of methods as Evolvability Indicators (EI), which are measures that attempt to capture how amendable the fitness landscape is to a GP search. The best examples of current EIs are the Fitness Distance Correlation (FDC) and the Negative Slope Coefficient (NSC). The second, more recent, group of methods are what we call Predictors of Expected Performance (PEP), which are predictive models that take as input a set of descriptive attributes of a particular problem and produce as output the expected performance of a GP system. The experimental work presented here compares an EI, the NSC, and a PEP model for a GP system applied to data classification. Results suggest that the EI fails at measuring problem difficulty expressed by the performance of the GP classifiers, an unexpected result. On the other hand, the PEP models show a very high correlation with the actual performance of the GP system. It appears that while an EI can correctly estimate the difficulty of a given search, as shown by previous research on this topic, it does not necessarily capture the difficulty of the underlying problem that GP is intended to solve. Conversely, while the PEP models treat the GP system as a computational black-box, they can still provide accurate performance predictions.

In [32], the goal is to predict the alertness of an individual by analyzing the brain activity through electroencephalographic data (EEG) captured with 58 electrodes. Alertness is characterized here as a binary variable that can be in a "normal" or "relaxed" state. We collected data from 44 subjects before and after a relaxation practice, giving a total of 88 records. After a pre-processing step and data validation, we analyzed each record and discriminate the alertness states using our proposed "slope criterion". Afterwards, several common methods for supervised classification (k nearest neighbors, decision trees (CART), random forests, PLS and discriminant sparse PLS) were applied as predictors for the state of alertness of each subject. The proposed "slope criterion" was further refined using a genetic algorithm to select the most important EEG electrodes in terms of classification accuracy. Results shown that the proposed strategy derives accurate predictive models of alertness.

6.5. Moderate Deviations for Mean Field Particle Models

The article [40] is concerned with moderate deviation principles of a general class of mean eld type interacting particle models. We discuss functional moderate deviations of the occupation measures for both the strong - topology on the space of fi nite and bounded measures as well as for the corresponding stochastic processes on some class of functions equipped with the uniform topology. Our approach is based on an original semigroup analysis combined with stochastic perturbation techniques and projective limit large deviation methods.

6.6. Bifurcating autoregressive processes

In [42], we investigate the asymptotic behavior of the least squares estimator of the unknown parameters of random coefficient bifurcating autoregressive processes. Under suitable assumptions on inherited and environmental effects, we establish the almost sure convergence of our estimates. In addition, we also prove a quadratic strong law and central limit theorems. Our approach mainly relies on asymptotic results for vector-valued martingales together with the well-known Rademacher-Menchov theorem.

In [46], we study the asymptotic behavior of the weighted least square estimators of the unknown parameters of random coefficient bifurcating autoregressive processes. Under suitable assumptions on the immigration and the inheritance, we establish the almost sure convergence of our estimators, as well as a quadratic strong law and central limit theorems. Our study mostly relies on limit theorems for vector-valued martingales.

In [47], we study the asymptotic behavior of the weighted least squares estimators of the unknown parameters of bifurcating integer-valued autoregressive processes. Under suitable assumptions on the immigration, we establish the almost sure convergence of our estimators, together with the quadratic strong law and central limit theorems. All our investigation relies on asymptotic results for vector-valued martingales.

6.7. Durbin-Watson statistic and first order autoregressive processes

In [45], we investigate moderate deviations for the Durbin-Watson statistic associated with the stable first-order autoregressive process where the driven noise is also given by a first-order autoregressive process. We first establish a moderate deviation principle for both the least squares estimator of the unknown parameter of the autoregressive process as well as for the serial correlation estimator associated with the driven noise. It enables us to provide a moderate deviation principle for the Durbin-Watson statistic in the easy case where the driven noise is normally distributed and in the more general case where the driven noise satisfies a less restrictive Chen-Ledoux type condition.

In [51], we investigate the asymptotic behavior of the Durbin-Watson statistic for the general stable $p-$order autoregressive process when the driven noise is given by a first-order autoregressive process. We establish the almost sure convergence and the asymptotic normality for both the least squares estimator of the unknown vector parameter of the autoregressive process as well as for the serial correlation estimator associated with the driven noise. In addition, the almost sure rates of convergence of our estimates are also provided. Then, we prove the almost sure convergence and the asymptotic normality for the Durbin-Watson statistic. Finally, we propose a new bilateral statistical procedure for testing the presence of a significative first-order residual autocorrelation and we also explain how our procedure performs better than the commonly used Box-Pierce and Ljung-Box statistical tests for white noise applied to the stable autoregressive process, even on small-sized samples.

6.8. Markovian superquadratic BSDEs

In [Stochastc Process. Appl., 122(9):3173-3208], the author proved the existence and the uniqueness of solutions to Markovian superquadratic BSDEs with an unbounded terminal condition when the generator and the terminal condition are locally Lipschitz. In [50], we prove that the existence result remains true for these BSDEs when the regularity assumptions on the generator and/or the terminal condition are weakened.

6.9. Non-Asymptotic Analysis of Adaptive and Annealed Feynman-Kac Particle Models

Sequential and Quantum Monte Carlo methods, as well as genetic type search algorithms can be interpreted as a mean field and interacting particle approximations of Feynman-Kac models in distribution spaces. The performance of these population Monte Carlo algorithms is strongly related to the stability properties of nonlinear Feynman-Kac semigroups. In [49], we analyze these models in terms of Dobrushin ergodic coefficients of the reference Markov transitions and the oscillations of the potential functions. Sufficient conditions for uniform concentration inequalities w.r.t. time are expressed explicitly in terms of these two quantities. We provide an original perturbation analysis that applies to annealed and adaptive FK models, yielding what seems to be the first results of this kind for these type of models. Special attention is devoted to the particular case of Boltzmann-Gibbs measures' sampling. In this context, we design an explicit way of tuning the number of Markov Chain Monte Carlo iterations with temperature schedule. We also propose and analyze an alternative interacting particle method based on an adaptive strategy to define the temperature increments.

6.10. A Robbins-Monro procedure for a class of models of deformation

The paper [48] deals with the statistical analysis of several data sets as- sociated with shape invariant models with different translation, height and scaling parameters. We propose to estimate these parameters together with the common shape function. Our approach extends the recent work of Bercu and Fraysse to multivariate shape invariant models. We propose a very efficient Robbins-Monro procedure for the estimation of the translation parameters and we use these esti- mates in order to evaluate scale parameters. The main pattern is estimated by a weighted Nadaraya-Watson estimator. We provide almost sure convergence and asymptotic normality for all estimators. Finally, we illustrate the convergence of our estimation procedure on simulated data as well as on real ECG data.

6.11. Individual load curves intraday forecasting

A dynamic coupled modelling is investigated to take temperature into account in the individual energy consumption forecasting. The objective in [44] is both to avoid the inherent complexity of exhaustive SARIMAX models and to take advantage of the usual linear relation between energy consumption and temperature for thermosensitive customers. We first recall some issues related to individual load curves forecasting. Then, we propose and study the properties of a dynamic coupled modelling taking temperature into account as an exogenous contribution and its application to the intraday prediction of energy consumption. Finally, these theoretical results are illustrated on a real individual load curve. The authors discuss the relevance of such an approach and anticipate that it could form a substantial alternative to the commonly used methods for energy consumption forecasting of individual customers.

APICS Project-Team

6. New Results

6.1. Source recovery problems

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Ana-Maria Nicu.

The works presented here are done in collaboration with Maureen Clerc and Théo Papadopoulo from the Athena EPI, with Doug Hardin and Edward Saff from Vanderbilt University (Nashville, USA), and with Abderrazek Karoui (Univ. Bizerte, Tunisie) and Jean-Paul Marmorat (Centre de mathématiques appliquées (CMA), École des Mines de Paris).

This section in dedicated to inverse problems for 3-D Poisson-Laplace equations. Though the geometrical settings differ in the 2 sections below, the characterization of silent sources (that give rise to a vanishing potential at measurement points) is a common problem to both which has been recently achieved, see [37],[29], [39].These are sums of (distributional) derivatives of Sobolev functions vanishing on the boundary.

6.1.1. Application to EEG

In 3-D, functional or clinical active regions in the cortex are often represented by pointwise sources that have to be localized from measurements on the scalp of a potential satisfying a Laplace equation (EEG, electroencephalography). In the work [4] it was shown how to proceed via best rational approximation on a sequence of 2-D disks cut along the inner sphere, for the case where there are at most 2 sources. A milestone in a long-haul research on the behaviour of poles of best rational approximants of fixed degree to functions with branch points has been reached this year [14], which shows that the technique carries over to finitely many sources (see section 4.2). In this connection, a dedicated software “FindSources3D” (see section 5.6) has been developed, in collaboration with the team Athena [16], [26].

Further, it appears that in the rational approximation step of these schemes, *multiple* poles possess a nice behaviour with respect to the branched singularities. This is due to the very basic physical assumptions on the model (for EEG data, one should consider *triple* poles). Though numerically observed in [16], there is no mathematical justification so far why these multiple poles have such strong accumulation properties, which remains an intriguing observation.

Issues of robust interpolation on the sphere from incomplete pointwise data are also under study in order to improve numerical accuracy of our reconstruction schemes. Spherical harmonics, Slepian bases and related special functions are of special interest (thesis of A.-M. Nicu [13], [67]), while other techniques should be considered as well.

Also, magnetic data from MEG (magneto-encephalography) will soon become available, which should enhance the accuracy of source recovery algorithms.

It turns out that discretization issues in geophysics can also be approached by these approximation techniques. Namely, in geodesy or for GPS computations, one may need to get a best discrete approximation of the gravitational potential on the Earth’s surface, from partial data collected there. This is the topic of a beginning collaboration with a physicist colleague (IGN, LAREG, geodesy). Related geometrical issues (finding out the geoid, level surface of the gravitational potential) are worthy of consideration as well.

6.1.2. Magnetization issues

Magnetic sources localization from observations of the field away from the support of the magnetization is an issue under investigation in a joint effort with the Math. department of Vanderbilt University and the Earth Sciences department at MIT. The goal is to recover the magnetic properties of rock samples (*e.g.* meteorites or stalactites) from fine field measurements close to the sample that can nowadays be obtained using SQUIDs (supraconducting coil devices).

The magnetization operator is the Riesz potential of the divergence of the magnetization. The problem of recovering a thin plate magnetization distribution from measurements of the field in a plane above the sample lead us to an analysis of the kernel of this operator, which we characterized in various function and distribution spaces (arbitrary compactly supported distributions or derivatives of bounded functions). For this purpose, we introduced a generalization of the Hodge decomposition in terms of Riesz transforms and showed that a thin plate magnetization is “silent” (i.e. in the kernel) if the normal component is zero and the tangential component is divergence free. In particular, we show that a unidirectional non-trivial magnetization with compact support cannot be silent. The same is true for bidirectional magnetizations if at least one of the directions is nontangential. We also proved that any magnetization is equivalent to a unidirectional. We did introduce notions of being silent from above and silent from below, which are in general distinct. These results have been reported in a paper to appear [37].

We currently work on Fourier based inversion techniques for unidirectional magnetizations, and Figures 5 , 6 , 7 and 8 show an example of reconstruction. A joint paper with our collaborators from VU and MIT is being written on this topic.

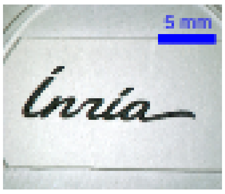


Figure 5. Inria’s logo were printed on a piece of paper. The ink of the letters “In” were magnetized along a direction D_1. The ink of the letters “ria” were magnetized along another direction D_2 (almost orthogonal to D_1).

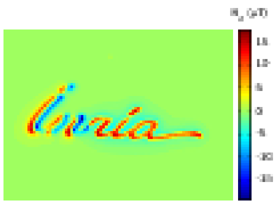


Figure 6. The Z-component of the magnetic field generated by the sample is measured by a SQUID microscope. The measure is performed 200µm above the sample.

For more general magnetizations, the severe ill-posedness of reconstruction challenges discrete Fourier methods, one of the main problems being the truncation of the observations outside the range of the SQUID measurements. We look forward to develop extrapolation techniques in the spirit of step 1 in section 3.1 .

6.2. Boundary value problems, generalized Hardy classes

Participants: Laurent Baratchart, Slah Chaabi, Juliette Leblond, Dmitry Ponomarev.

This work has been performed in collaboration with Yannick Fischer from the Magique3D EPI (Inria Bordeaux, Pau).

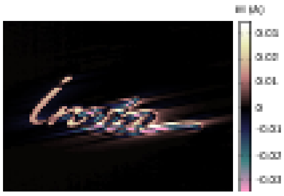


Figure 7. The field measured in Figure 6 is inversed, assuming that the sample is unidimensionally magnetized along the direction D_1. The letters "In" are fairly well recovered while the rest of the letters is blurred (because the hypothesis about the direction of magnetization is false for "ria").

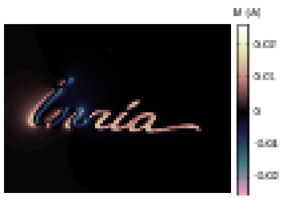


Figure 8. The field measured in Figure 6 is inversed, assuming that the sample is unidimensionally magnetized along the direction D_2. The letters "ria" are fairly well recovered while the rest of the letters is blured (because the hypothesis about the direction of magnetization is false for "In").

In collaboration with the CMI-LATP (University Aix-Marseille I), the team considers 2-D diffusion processes with variable conductivity. In particular its complexified version, the so-called *conjugate* or *real Beltrami equation*, was investigated. In the case of a smooth domain, and for Lipschitz conductivity, we analyzed the Dirichlet problem for solutions in Sobolev and then in Hardy classes [5].

Their traces merely lie in L^p $(1 < p < \infty)$ of the boundary, a space which is suitable for identification from pointwise measurements. Again these traces turn out to be dense on strict subsets of the boundary. This allows us to state Cauchy problems as bounded extremal issues in L^p classes of generalized analytic functions, in a reminiscent manner of what was done for analytic functions as discussed in section 3.3.1 .

We generalized the construction to finitely connected Dini-smooth domains and $W^{1,q}$-smooth conductivities, with $q > 2$ [35]. The case of an annular geometry is the relevant one for the application to plasma shaping mentioned below [58], [35]. The application that initially motivated this work came from free boundary problems in plasma confinement (in tokamaks) for thermonuclear fusion. This work was initiated in collaboration with the Laboratoire J. Dieudonné (University of Nice).

In the transversal section of a tokamak (which is a disk if the vessel is idealized into a torus), the so-called poloidal flux is subject to some conductivity equation outside the plasma volume for some simple explicit smooth conductivity function, while the boundary of the plasma (in the Tore Supra tokamak) is a level line of this flux [54]. Related magnetic measurements are available on the chamber, which furnish incomplete boundary data from which one wants to recover the inner (plasma) boundary. This free boundary problem (of Bernoulli type) can be handled through the solutions of a family of bounded extremal problems in generalized Hardy classes of solutions to real Beltrami equations, in the annular framework [35].

In the particular case at hand, the conductivity is $1/x$ and the domain is an annulus embedded in the right half-plane. We obtained a basis of solutions (exponentials times Legendre functions) upon separating variables in toroidal coordinates. This provides a computational setting to solve the extremal problems mentioned before, and was the topic of the PhD thesis of Y. Fischer [58], [27]. In the most recent tokamaks, like Jet or ITER, an interesting feature of the level curves of the poloidal flux is the occurrence of a cusp (a saddle point of the poloidal flux, called an X point), and it is desirable to shape the plasma according to a level line passing through this X point for physical reasons related to the efficiency of the energy transfer. We established well-posedness of the Dirichlet problem in weighted L^p classes for harmonic measure on piecewise smooth domains without cusps, thereby laying ground for such a study. This issue is next in line, now that the present approach has been validated numerically on Tore Supra data, and the topic of the PhD thesis of D. Ponomarev.

The PhD work of S. Chaabi is devoted to further aspects of Dirichlet problems for the conjugate Beltrami equation. On the one hand, a method based on Foka's approach to boundary value problems, which uses Lax pairs and solves for a Riemann-Hilbert problem, has been devised to compute in semi explicit form solutions to Dirichlet and Neumann problems for the conductivity equation satisfied by the poloidal flux. Also, for more general conductivities, namely bounded below and lying in $W^{1,s}$ with $s \geq 2$, parameterization of solutions to Dirichlet problems on the disk by Hardy function was achieved through Bers-Nirenberg factorization. Note the conductivity may be unbounded when $s = 2$, which is completely new. Two papers are being prepared reporting on these topics.

Finally, note that the conductivity equation can be expressed like a static Schrödinger equation, for smooth enough conductivity coefficients. This provides a link with the following results recently set up by D. Ponomarev, who recently join the team for his PhD. A description of laser beam propagation in photopolymers can be crudely approximated by a stationary two-dimensional model of wave propagation in a medium with negligible change of refractive index. In such setting, Helmholtz equation is approximated by a linear Schrödinger equation with one of spatial coordinates being an evolutionary variable. Explicit comparison of the solutions in the whole half-space allows to establish global justification of the Schrodinger model for sufficiently smooth pulses [73]. This phenomenon can also be described by a nonstationary model that relies on the spatial nonlinear Schrödinger (NLS) equation with the time-dependent refractive index. A toy problem is considered in [71], when the rate of change of refractive index is proportional to the squared amplitude of the electric field and the spatial domain is a plane. The NLS approximation is derived from a 2-D quasi-linear wave equation, for small time intervals and smooth initial data. Numerical simulations illustrate the approximation result in the 1-D case.

6.3. Circuit realisations of filter responses: determination of canonical forms and exhaustive computations of constrained realisations

Participant: Fabien Seyfert.

This work has been done in collaboration with Smain Amari (Royal Military College, Kingston, Canada), Jean Charles Faugère (SALSA EPI, Inria Rocquencourt), Giuseppe Macchiarella (Politecnico di Milano, Milan, Italy), Uwe Rosenberg (Design and Project Engineering, Osterholz-Scharmbeck, Germany) and Matteo Oldoni (Politecnico di Milano, Milan, Italy).

We continued our work on the circuit realizations of filters' responses with mixed type (inductive or capacitive) coupling elements and constrained topologies [1]. For inline circuits, methods based on sequential extractions of electrical elements are best suited due to their computational simplicity. On the other hand, for circuits with no inline topology ,such methods are inefficient while algebraic methods (based on a Groebner basis) can be used, but at high computational cost. In order to tackle large order circuits, our approach is to decompose them into connected inline sections, which can be directly realized by extraction techniques, and into complex sections, where algebraic methods are needed for realization. In order to do this, we started studying the synthesis of filter responses by means of circuits with reactive non-resonating nodes (dangling resonators) [22]. Links of this topic with Potapov's factorization of J-inner functions are currently being investigated.

In this connection, sensitivity analysis of the electrical response of a filter with respect to the electrical parameters of the underlying circuit has been published in collaboration with the University of Cartagena and ESA [20]. We essentially proved that the total electrical sensitivity of a filters' response does not depend on the coupling topology of the underlying circuit: the latter however controls the distribution of this sensitivity within each resonator.

6.4. Synthesis of compact multiplexers and de-embedding of multiplexers

Participants: Martine Olivi, Sanda Lefteriu, Fabien Seyfert.

This work has been done in collaboration with Stéphane Bila (Xlim, Limoges, France), Hussein Ezzedin (Xlim, Limoges, France), Damien Pacaud (Thales Alenia Space, Toulouse, France), Giuseppe Macchiarella (Politecnico di Milano, Milan, Italy, and Matteo Oldoni (Politecnico di Milano, Milan, Italy).

6.4.1. *Synthesis of compact multiplexers*

We focused our research on multiplexer with a star topology. These are comprised of a central N-port junction, and of filters plugged on all but common ports (see Figure 9). A possible approach to synthesis of the multiplexer's response is to postulate that each filter channel has to match the multiplexer at n_k frequencies (n_k being the order of the filter) while rejecting the energy at m_k other frequencies (m_k being the order the transmission polynomial of the filter). The desired synthesis can then be cast into computing of a collection of filter's responses matching the energy as prescribed and rejecting it at specified frequencies when plugged simultaneously on the junction. Whether such a collection exists is one of the main open issues facing co-integration of systems in electronics. Investigating the latter led us to consider the simpler problem of matching a filter, on a frequency-varying load, while rejecting energy at fixed specified frequencies. If the order of the filter is n this amounts to fix a given transmission polynomial r and to solve for a unitary polynomial p meeting interpolation conditions of the form:

$$j = 1..n, \qquad \frac{p}{q}(w_j) = \gamma_j$$

where q is the unique monic Hurwitz polynomial satisfying the Feldtkeller equation

$$qq^* = pp^* + rr^*.$$

This problem can be seen as an extended Nevanlinna-Pick interpolation problem, which was considered in [62] when the interpolation frequencies lie in the *open* left half-plane. We conjecture that existence and uniqueness of the solution still holds in our case, where interpolation takes place on the boundary, provided r has no roots on the imaginary axis. Numerical experiments based on continuation techniques tend to corroborate our belief: efforts now focus on a mathematical proof. The derived numerical tools have already been used to successfully to design multiplexer's responses in collaboration with CNES and Xlim, thereby initiating a collaboration with Xlim on co-integration of filters and antennas.

6.4.2. De-embedding of multiplexers

Let S be the measured scattering matrix of a multiplexer composed of a N-port junction with response T and $N-1$ filters with responses $F_1, \cdots F_{N-1}$ as plotted on Figure 9 . The de-embedding question we raise is the following: given S and T, is it possible to retrieve the F_k's ? The answer to this question depends of course of the admissible class of filters. For the simplest case where no assumption is made (except reciprocity), we showed that the problem has a unique solution for $N > 3$ and for generic T, while for $N = 2$ the solution space at each frequency point has real dimension 2. This redundancy can be explained by the existence of "ghost" or "silent" components that can hide behind the junction: when being chained to the junction these components do not affect its response. We also experienced that the generic behaviour for $N > 3$ is rather theoretical, as usual junctions are often made of chained T-junctions: in this non generic case (which is rather generic in practice !) some "silent" component still exists for $N > 3$. Additional hypotheses, such as rationality with prescribed degree for F_k, are currently being studied and already yielded results for the case $N = 3$ [21].

This work is pursued in collaboration with Thales Alenia Space, Politecnico di Milano, Xlim and CNES in particular within the contract CNES-Inria on compact N-port synthesis (see section 7.1).

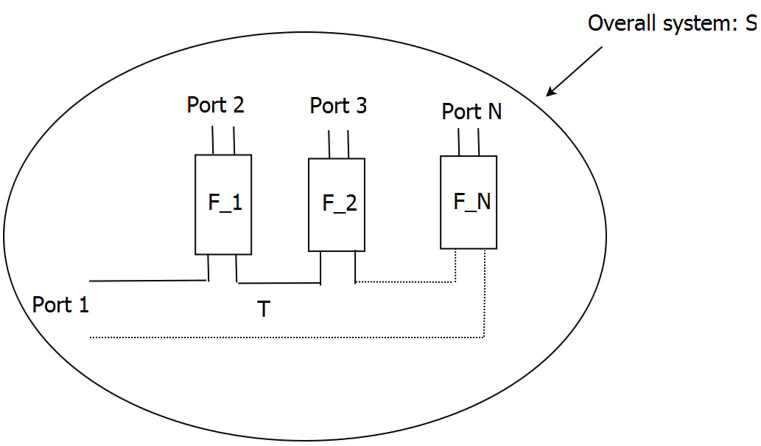


Figure 9. Multiplexer made of a junction T and filtering devices $F_1, F_2 \cdots F_N$

6.5. Detection of the instability of amplifiers

Participants: Laurent Baratchart, Sylvain Chevillard, Martine Olivi, Fabien Seyfert.

This work is conducted in collaboration with Jean-Baptiste Pomet from the McTao team. It is a continuation of a collaboration with CNES and the University of Bilbao.The goal is to help developing amplifiers, in particular to detect instability at an early stage of the design.

Currently, Electrical Engineers from the University of Bilbao, under contract with CNES (the French Space Agency), use heuristics to diagnose instability before the circuit is physically implemented. We intend to set up a rigorously founded algorithm, based on properties of transfer functions of such amplifiers which belong to particular classes of analytic functions.

In non-degenerate cases, non-linear electrical components can be replaced by their first order approximation when studying stability to small perturbations. Using this approximation, diodes appear as perfect negative resistors and transistors as perfect current sources controlled by the voltages at certain points of the circuit.

In 2011, we had proved that the class of transfer functions which can be realized with such ideal components and standard passive components (resistors, selfs, capacitors and transmission lines) is rather large since it contains all rational functions in the variable and in the exponentials thereof.

In 2012, we focused on the kind of instabilities that these ideal systems can exhibit. We showed that a circuit can be unstable, although it has no pole in the right half-plane. This remains true even if a high resistor is put in parallel of the circuit, which is rather unusual. This pathological example is unrealistic, though, because it assumes that non-linear elements continue to provide gain even at very high frequencies. In practice, small capacitive and inductive effects (negligible at moderate frequencies) make these components passive for very high frequencies. Under this simple assumption, we proved that the class of transfer functions of realistic circuits is much smaller than in previous situation. In fact, a realistic circuit is unstable if and only if it has poles in the right half-plane. Moreover, there can only be finitely many of them. An article is currently being written on the subject.

6.6. Best constrained analytic approximation

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Dmitry Ponomarev, Elodie Pozzi.

This work is performed in collaboration with Jonathan Partington (Univ. Leeds, UK).

Continuing effort is being paid by the team to carry over the solution to bounded extremal problems of section 3.3.1 to various settings. We mentioned already in section 6.2 the extension to 2-D diffusion equations with variable conductivity for the determination of free boundaries in plasma control and the development of a generalized Hardy class theory. We also investigate the ordinary Laplacian in $\mathbb{R}^3$, where targeted applications are to data transmission step for source detection in electro/magneto-encephalography (EEG/MEG, see section 6.1).

Still, questions about the behaviour of solutions to the standard bounded extremal problems (P) of section 3.3.1 deserve attention. We realized this year that Slepian functions are eigenfunctions of truncated Toeplitz operators in 2-D. This can be used to quantify robustness properties of our resolution schemes in H^2 and to establish error estimates, see [25]. Moreover we considered additional interpolation constraints [28], as a simplified but already interesting issue, before getting at extremal problems for generalized analytic functions in annular non-smooth domains. The latter arise in the context of plasma shaping in tokamaks like ITER, and will be the subject of the PhD thesis of D. Ponomarev.

In another connection, weighted composition operators on Lebesgue, Sobolev, and Hardy spaces appear in changes of variables while expressing conformal equivalence of plane domains. A universality property related to the existence of invariant subspaces for these important classes of operators has been established in [19]. Additional density properties also allow one to handle some of their dynamical aspects (like cyclicity).

6.7. Rational Approximation for fitting Non-Negative EPT densities

Participants: Martine Olivi, Fabien Seyfert.

This work has been done in collaboration with Bernard Hanzon and Conor Sexton from Univ. Cork.

The problem is to fit a probability density function on a large set of financial data. The model class is the set of non-negative EPT (Exponential-Polynomials-Trigonometric) functions which provides a useful framework for probabilistic calculation as illustrated in the link http://www.2-ept.com/2-ept-literature.html. Moreover, an EPT function can alternatively be interpreted as the impulse response of a continuous time stable system whose Laplace transform is a rational transfer function. This interpretation allows us to approach this problem using approximation tools developed by the team. The very context brings up a classical, as yet essentially unsolved difficulty in rational approximation, namely preservation of positivity. This is known to be a hard issue. Our work, initiated in 2011, resulted this year in an improved approach for checking non-negativity of an EPT function. These results have been presented at the 16th IFAC Conference on System Identification [23]. The proposed method was demonstrated on the positive daily Dow Jones Industrial Average (DJIA) log returns over 80 years.

6.8. Rational and meromorphic approximation

Participant: Laurent Baratchart.

This work has been done in collaboration with Herbert Stahl (TFH Berlin) and Maxim Yattselev (Univ. Oregon at Eugene, USA).

We completed and published this year the proof of an important result in approximation theory, namely the counting measure of poles of best H^2 approximants (more generally: of critical points) of degree n to a function analytically continuable, except over finitely many branchpoints lying outside the unit disk, converges to the Green equilibrium distribution of the compact set of minimal Green capacity outside of which the function is single valued [14]. The proof requires showing existence and uniqueness of a compact set of minimal weighted logarithmic capacity in a field, outside of which the function is single-valued. Structure of this contour, along with error estimates, also come out of the proof. The result is in fact valid for functions that are Cauchy integrals of Dini-smooth functions on such a contour. We rely in addition on asymptotic interpolation estimates from [63].

This result warrants source recovery techniques used in section 6.1.1 .

We also studied partial realizations, or equivalently Padé approximants to transfer functions with branchpoints. Identification techniques based on partial realizations of a stable infinite-dimensional transfer function are known to often provide unstable models, but the question as to whether this is due to noise or to intrinsic instability was not clear. In the case of 4 branchpoints, expressing the computation of Padé approximants in terms of the solution to a Riemann-Hilbert problem on the Riemann surface of the function, we proved that the pole behaviour generically shows deterministic chaos [49].

6.9. Tools for numerically guaranteed computations

Participant: Sylvain Chevillard.

The overall and long-term goal is to enhance the quality of numerical computations. The progress made during year 2012 is the following:

- Publication of a work about the implementation of functions erf and erfc in multiple precision and with correct rounding [15]. It corresponds to a work initially begun in the Arénaire team and finished in the Caramel team. The goal of this work is to show on a representative example the different steps of the rigorous implementation of a function in multiple precision arithmetic (choice of a series approximating the function, choice of the truncation rank and working precision used for the computation, roundoff analysis, etc.). The steps are described in such a way that they can easily be reproduced by someone who would like to implement another function. Moreover, it is showed that the process is very regular, which suggests that it (or at least large parts of it) could be automated.
- In the same field of multiple precision arithmetic, and with Marc Mezzarobba (Aric team), we proposed an algorithm for the efficient evaluation of the Airy $\mathrm{Ai}(x)$ function when x is moderately large [57]. Again, this work deals with a representative example, with the idea of trying to automate

the process as a future work. The Taylor series of the Airy Ai function (as many others such as, e.g., Bessel functions or erf) is ill-conditioned when x is not small. To overcome this difficulty, we extend a method by Gawronski, Müller and Reinhard, known to solve the issue in the case of the error function erf. We rewrite $\mathrm{Ai}(x)$ as $G(x)/F(x)$ where F and G are two functions with well-conditioned series. However, the coefficients of G turn out to obey a three-terms ill-conditioned recurrence. We evaluate this recurrence using Miller's backward algorithm with a rigorous error analysis.

- Finally, a more general endeavor is to develop a tool that helps developers of libms in their task. This is performed by the software Sollya [2], developed in collaboration with C. Lauter (Université Pierre et Marie Curie) and M. Joldeş (Uppsala University). During year 2012, a large effort has been made in view of the release of version 4.0 (to come in 2013). This effort (of about 400 commits in the svn repository of the project) is mainly intended to provide a library version of Sollya, as well as a robust test suite for the tool. As a matter of course, it allowed us to detect and fix a dozen of bugs.

[2] http://sollya.gforge.inria.fr/

ASPI Project-Team

5. New Results

5.1. On the length of one–dimensional reactive paths

Participants: Frédéric Cérou, Arnaud Guyader, Florent Malrieu.

See 3.3 and 4.2 .

This is a collaboration with Tony Lelièvre (ENPC).

Motivated by some numerical observations on molecular dynamics simulations, we analyze metastable trajectories in a very simple setting, namely paths generated by a one-dimensional overdamped Langevin equation for a double well potential. More precisely, we are interested in so–called reactive paths, namely trajectories which leave definitely one well and reach the other one. The aim of [32] is to precisely analyze the distribution of the lengths of reactive paths in the limit of small temperature, and to compare the theoretical results to numerical results obtained by a Monte Carlo method, namely the multi–level splitting approach.

5.2. Long time behavior of piecewise–deterministic Markov processes

Participant: Florent Malrieu.

This is a collaboration with Michel Benaïm (université de Neuchâtel), Stéphane Le Borgne (IRMAR) and Pierre–André Zitt (université de Marne–la–Vallée).

5.2.1. Quantitative ergodicity for some switched dynamical systems

We provide quantitative bounds for the long time behavior of a class of piecewise deterministic Markov processes with state space $R^d \times E$ where E is a finite set. The continuous component evolves according to a smooth vector field that switches at the jump times of the discrete coordinate. The jump rates may depend on the whole position of the process. Under regularity assumptions on the jump rates and stability conditions for the vector fields we provide explicit exponential upper bounds for the convergence to equilibrium in terms of Wasserstein distances [13]. As an example, we obtain convergence results for a stochastic version of the Morris–Lecar model of neurobiology.

5.2.2. On the stability of planar randomly switched systems

Consider the random process (X_t) solution of $dX_t/dt = A(I_t)X_t$ where (I_t) is a Markov process on $\{0,1\}$ and A_0 and A_1 are real Hurwitz matrices on R^2. Assuming that there exists $\lambda \in (0,1)$ such that $(1-\lambda)A_0 + \lambda A_1$ has a positive eigenvalue, we establish that the norm of X_t may converge to 0 or infinity, depending on the the jump rate of the process I. An application to product of random matrices is studied. The paper [29] can be viewed as a probabilistic counterpart of the paper [36] by Baldé, Boscain and Mason.

5.2.3. Qualitative properties of certain piecewise deterministic Markov processes

We study a class of piecewise deterministic Markov processes with state space $R^m \times E$ where E is a finite set. The continous component evolves according to a smooth vector field that it switched at the jump times of the discrete coordinate. The jump rates may depend on the whole position of the process. Working under the general assumption that the process stays in a compact set, we detail a possible construction of the process and characterize its support, in terms of the solutions set of a differential inclusion. We establish results on the long time behaviour of the process, in relation to a certain set of accessible points, which is shown to be strongly linked to the support of invariant measures. Under Hörmander–type bracket conditions, we prove that there exists a unique invariant measure and that the processes converges to equilibrium in total variation. Finally we give examples where the bracket condition does not hold, and where there may be one or many invariant measures, depending on the jump rates between the flows [30].

5.3. Quantitative long time behavior of an ergodic variant of the telegraph process

Participant: Florent Malrieu.

This is a collaboration with Joaquin Fontbona (University of Chile) and Hélène Guérin (IRMAR).

Motivated by stability questions on piecewise deterministic Markov models of bacterial chemotaxis, we study the long time behavior of a variant of the classic telegraph process having a non–constant jump rate that induces a drift towards the origin. We compute its invariant law and show exponential ergodicity, obtaining a quantitative control of the total variation distance to equilibrium at each instant of time. These results [15] rely on an exact description of the excursions of the process away from the origin and on the explicit construction of an original coalescent coupling for both velocity and position. Sharpness of the obtained convergence rate is discussed.

5.4. Total variation estimates for the TCP process

Participant: Florent Malrieu.

This is a collaboration with Jean-Baptiste Bardet (université de Rouen), Alejandra Christen (University of Chile), Arnaud Guillin (université de Clermont–Ferrand), and Pierre–André Zitt (université de Marne–la–Vallée).

The TCP window size process appears in the modeling of the famous Transmission Control Protocol used for data transmission over the Internet. This continuous time Markov process takes its values in $[0, \infty)$, is ergodic and irreversible. The sample paths are piecewise linear deterministic and the whole randomness of the dynamics comes from the jump mechanism. The aim of [28] is to provide quantitative estimates for the exponential convergence to equilibrium, in terms of the total variation and Wasserstein distances.

5.5. Convergence results for approximate Bayesian computation

Participants: Frédéric Cérou, Arnaud Guyader.

This is a collaboration with Gérard Biau (ENS and université Pierre et Marie Curie).

Approximate Bayesian computation (ABC for short) is a family of computational techniques which offer an almost automated solution in situations where evaluation of the posterior likelihood is computationally prohibitive, or whenever suitable likelihoods are not available. In [31], we analyze the procedure from the point of view of k-nearest neighbor theory and explore the statistical properties of its outputs. We discuss in particular some asymptotic features of the genuine conditional density estimate associated with ABC, which is a new interesting hybrid between a k-nearest neighbor and a kernel method. These are among the very few results on the convergence of ABC, and our assumptions on the underlying probability distribution are minimal.

5.6. Soft level splitting for rare event estimation

Participants: Frédéric Cérou, Arnaud Guyader.

See 3.3 and 4.2 .

This is a collaboration with Nicolas Hengartner (Los Alamos).

It is well established now that one can use adaptive splitting levels to compute the conditional probabilities of nested sets. To get an efficient algorithm, the probability of a set given the previous one should be always the same, which is approximately achieved adaptively by using the empirical cdf (cumulative distribution function) of the scores. The way to proceed is to fix a probability of success p_0, and then choose the p_0 quantile of the current scores. Here we investigate whether, by using the whole cdf, and not only one quantile, we can design an algorithm with better performance. The main trick is a transformation to have a sample of exponential variables. This would require the knowledge of the cdf of the cost, which is obviously unvailable, but we can replace it by the empirical cdf of the sample at the previous level. The complete theoretical study of this algorithm is still to be done, but we have illustrated by some examples that it can lead to significantly better results than the standard splitting procedure with the same number of intermediate levels.

5.7. Decoding fingerprints using the Markov chain Monte Carlo method

Participants: Frédéric Cérou, Arnaud Guyader.

This is a collaboration with Teddy Furon (Inria Rennes, project–team TEXMEX).

The paper [22] proposes a new fingerprinting decoder based on the Markov chain Monte Carlo (MCMC) method. A Gibbs sampler generates groups of users according to the posterior probability that these users could have forged the sequence extracted from the pirated content. The marginal probability that a given user pertains to the collusion is then estimated by a Monte Carlo method. The users having the biggest empirical marginal probabilities are accused. This MCMC method can decode any type of fingerprinting codes. This paper is in the spirit of the *learn and match* decoding strategy: it assumes that the collusion attack belongs to a family of models. The expectation–maximization algorithm estimates the parameters of the collusion model from the extracted sequence. This part of the algorithm is described for the binary Tardos code and with the exploitation of the soft outputs of the watermarking decoder. The experimental body considers some extreme setups where the fingerprinting code lengths are very small. It reveals that the weak link of our approach is the estimation part. This is a clear warning to the *learn and match* decoding strategy.

5.8. Iterative isotone regression

Participants: Arnaud Guyader, Nicolas Jégou.

This is a collaboration with Nicolas Hengartner (Los Alamos) and Eric Matzner–Løber (université de Rennes 2), and with Alexander B. Németh (Babeş Bolyai University) and Sándor Z. Németh (University of Birmingham).

The current collaboration on nonparametric regression focuses on a novel nonparametric regression technique that applies ideas borrowed from iterative bias reduction to estimating functions of bounded variations. This work has emerged from the joint supervision of Nicolas Jégou's PhD thesis by Arnaud Guyader, Nick Hengartner and Eric Matzner-Løber.

A geometric approach has been investigated, as an extension of some ideas developed in the thesis. The current work [33] proposes and analyzes a novel method for estimating a univariate regression function of bounded variation. The underpinning idea is to combine two classical tools in nonparametric statistics, namely isotonic regression and the estimation of additive models. A geometrical interpretation enables us to link this iterative method with Von Neumann's algorithm. Moreover, making a connection with the general property of isotonicity of projection onto convex cones, we derive another equivalent algorithm and go further in the analysis. As iterating the algorithm leads to overfitting, several practical stopping criteria are also presented and discussed.

5.9. Detection issues in track–before–detect

Participants: François Le Gland, Alexandre Lepoutre.

See 4.1 .

This is a collaboration with Olivier Rabaste (ONERA Palaiseau).

Track–before–detect refers to situations where the target SNR is so low that it is practically impossible to detect the presence of a target, using a simple thresholding rule. In such situations, the solution is to keep all the information available in the raw radar data and to address directly the tracking problem, using a particle filter with a binary Markov variable that models the presence or absence of the target. The choice of the proposal distribution is crucial here, and an efficient particle filter is proposed [24] that is based on a relevant proposal distribution built from detection and estimation considerations, that aims at extracting all the available information from the measurements. The proposed filter leads to a dramatically improved performance as compared with particle filters based on the classical proposal distribution, both in terms of detection and estimation. A further improvement, in terms of detection performance, is to model the problem as a quickest change detection problem [70] in a Bayesian framework. In this context, the posterior distribution of the first time of appearance of the target is a mixture where each component represents the hypothesis that the target appeared at a given time. The posterior distribution is intractable in practice, and it is proposed [23] to approximate each component of the mixture by a particle filter. It turns out that the mixture weights can be computed recursively in terms of quantities that are provided by the different particle filters. The overall filter yields good performance as compared with classical particle filters for track–before–detect.

5.10. Estimation of conflict probability

Participants: François Le Gland, Damien–Barthélémy Jacquemart.

See 3.3 and 4.2 .

This is a collaboration with Jérôme Morio (ONERA Palaiseau).

In [16], the conflict probability between aircraft in uncontrolled airspace is estimated using the importance splitting method, and this algorithm is applied on realistic situations of aircraft conflict. The current work aims at designing efficient intermediate regions at a reasonnable computational cost, or alternatively at introducing weights to compensate for a simple but suboptimal design of the intermediate regions.

5.11. Minimum volume set for a rare event

Participants: François Le Gland, Rudy Pastel.

See 3.3 and 4.2 .

This is a collaboration with Jérôme Morio (ONERA Palaiseau).

The paper [19] first reviews the principle of minimum volume set estimation of a given probability level for a multidimensional density, a strategy that provides a sound solution to the multidimensional quantile issue. It then describes an importance sampling algorithm that is suitable for this kind of estimation problems, and provides simulation results for the estimation of the impact zone of a space launcher. The current work aims at designing an importance splitting method that would be more efficient for extreme quantiles.

5.12. Laplace and sequential Monte Carlo methods in Bayesian filtering

Participants: François Le Gland, Paul Bui–Quang.

This is a collaboration with Christian Musso (ONERA Palaiseau).

The Laplace method is a deterministic technique to approximate integrals, and it has been widely used in Bayesian statistics, e.g. to compute posterior means and variances [72]. The approximation is consistent as the observations sample size goes to infinity or as the observation noise intensity goes to zero, and the main condition to apply the method is that the model should be identifiable. The aim of [21] is to combine SMC methods and the Laplace method in order to better approximate the posterior density in nonlinear Bayesian filtering. At each stage of the proposed algorithm, a first approximate density is build from the current population of particles, then an accurate estimate of the posterior mean and covariance matrix is obtained using the Laplace method, and these estimates are used to shift and rescale the population of particles. Overall, this procedure could be interpreted as another design of an importance distribution that takes the observations into account. The current work aims at using the Laplace method to cope with *weight degeneracy* in particle filtering, a phenomenon that typically occurs when the observation noise is small, which is precisely the situation where the Laplace method is efficient.

5.13. Wind–wave modelling

Participant: Valérie Monbet.

This is a collaboration with Pierre Ailliot (UBO).

Climate change will bring large changes to the mean climate, and especially to climate extremes, over the coming decades. Computationally expensive global climate model (GCM) projections provide good information about future mean changes. Computationally efficient, yet physically consistent, statistical models of weather variables (stochastic weather generators) allow us to explore the frequency and severity of weather and climate events in much greater detail. When deployed as a complement to GCMs, stochastic weather generators provide a much richer picture of the future, allowing us to better understand, evaluate and manage future weather and climate risks, especially for renewal energy. In this context we are developing a space time model for wind fields in the North–East Atlantic, based on a conditionally transformed Gaussian state space model.

5.14. Sequential data assimilation: ensemble Kalman filter vs. particle filter

Participants: François Le Gland, Valérie Monbet.

Surprisingly, very little was known about the asymptotic behaviour of the ensemble Kalman filter [44], [45], [46], whereas on the other hand, the asymptotic behaviour of many different classes of particle filters is well understood, as the number of particles goes to infinity. Interpreting the ensemble elements as a population of particles with mean–field interactions, and not only as an instrumental device producing an estimation of the hidden state as the ensemble mean value, it has been possible to prove the convergence of the ensemble Kalman filter, with a rate of order $1/\sqrt{N}$, as the number N of ensemble elements increases to infinity [62]. In addition, the limit of the empirical distribution of the ensemble elements has been exhibited, which differs from the usual Bayesian filter. The next step has been to prove (by induction) the asymptotic normality of the estimation error, i.e. to prove a central limit theorem for the ensemble Kalman filter.

BACCHUS Team

6. New Results

6.1. Residual distribution schemes

Participants: Rémi Abgrall [Corresponding member], Mario Ricchiuto, Dante De Santis, Algiane Froehly, Cécile Dobrzynski.

We have understood how to approximate the advection diffusion problem in the context of residual distribution schemes. A third order version for scalar problem has been written. It is uniformly accurate, from pure viscous to pure convection problems. This scheme has been generalised to the laminar Navier Stokes equations. An extension to the turbulent case (with Spalart Allmaras model) has also been written and tested. We have studied the (iterative) convergence issues using Jacobian Free techniques or the LUSGS algorithm. Tests in two and three dimensions have been carried out. This work is submitted in [37] and has been the topic of [20].

A. Froehly has submitted her PhD thesis about the extension of the residual distribution scheme using isogeometric analysis. In particular, we have foccussed on mesh adaption, including at the boundary. A paper is being written to summarized the work.

6.2. Curved meshes

Participants: Rémi Abgrall, Cécile Dobrzynski [Corresponding member], Algiane Froehly.

One of the main open problems in high order schemes is the design of meshes that fit with enough accuracy the boundary of the computational domain. If this curve/surface is not locally straight/planar, the elements must be curved near the boundary, and their curvature need to be propagated to the interior of the domain to have valid elements. When the mesh is very streched, this can be quite challenging since, in addition, we want that the mesh keep a structure, in particular for boundary layers. Using tools explored in isogeometrical analysis, we have been able to construct a prototype computing curved meshes (in 2D and 3D), while keeping the structure of the mesh.

6.3. Hypoelastic models

Participants: Rémi Abgrall [Corresponding member], Pierre-Henri Maire.

In collaboration with CEA (P.H. maire), we have developped and tested a new finite volume like algorithm able to simulate hypoelastic-plastics problems on unstructured meshes. This has been published in [47].

6.4. Penalisation methods using unstructured meshes

Participants: Rémi Abgrall, Cécile Dobrzynski, Héloïse Beaugendre [Corresponding member].

In Computational Fluid Dynamics the interest on embedded boundary methods for Navier-Stokes equations increases because they simplify the meshing issue, the simulation of multi-physics flows and the coupling of fluid-solid interactions in situation of large motions or deformations. Nevertheless, an accurate treatment of the wall boundary conditions remains an issue of these methods. In this work we develop an immersed boundary method for unstructured meshes based on a penalization technique and we use mesh adaption to improve the accuracy of the method close to the boundary. The idea is to combine the strength of mesh adaptation, that is to provide an accurate flow description especially when dealing with wall boundary conditions, to the simplicity of embedded grids techniques, that is to simplify the meshing issue and the wall boundary treatment when combined with a penalization term to enforce boundary conditions. The bodies are described using a level-set method and are embedded in an unstructured grid. Once a first numerical solution is computed mesh adaptation based on two criteria the level-set and the quality of the solution is performed.

6.5. Unsteady problem

Participants: Rémi Abgrall, Mario Ricchiuto [Corresponding member], Luca Arpaia, Jan Klosa.

Using a reinterpretation of the explicit RD scheme we had designed 2 years ago, we have been able to construct a third order accurate RD scheme in one space dimension. The extension to multidimensional problems is pending.

We have studied the extention of second order unsteady RD scheme to the ALE formulation. New version of the explsiict unsteady RD schemes have been studied.

6.6. Lagrangian hydrodynamics

Participants: Rémi Abgrall [Corresponding member], Pierre-Henri Maire, François Vilar.

F. Vilar has achieved his thesis on the approximation of the Euler equations written in pure Lagrangian coordinates. He has foccussed on third order accuracy in time and space, usning a Discontinuous Galerkin formulation. The solution is approximated localy by quadratic polynomials. The boundary of elements are approximated by Bezier curves. He has managed to achieve an approximation consistant with the geometric Cosnervation Law. Many test cases have been computed, showing both a dramatic improvement of the accuracy and the robustness of the method with respect to its second order counterpart.

6.7. Boundary Layer Enrichment

Participants: Rémi Abgrall [Corresponding member], Arnaud Krust.

Arnaud Krust has finished his PhD thesis on boundary layer enrichment. We developed a numerical framework well suited for advection- diffusion problems when the advection part is dominant. In that case, given Dirichlet type boundary condition, it is well known that a boundary layer develops. In order to resolve correctly this layer, standard methods consist in increasing the mesh resolution and possibly increasing the formal accuracy of the numerical method. In this work, we follow another path: we do not seek to increase the formal accuracy of the scheme but, by a careful choice of finite element, to lower the mesh resolution in the layer. Indeed the finite element representation we choose is locally the sum of a standard one plus an enrichment. This work proposes such a method and with several numerical examples, we show the potential of this approach. In particular we show that the method is not very sensitive to the choice of the enrichment functions. The best choice of enrichment are shown to be obtained by an iterative mechanisms which bears some common features with mesh refinement.

6.8. Uncertainty Quantification

Participants: Rémi Abgrall, Pietro Congedo [Corresponding member], Gianluca Geraci, Mario Ricchiuto.

We developed two research lines: the first one focused on the computation of high-order statistics, the second one is related to the formulation of a global framework in the coupled physical/stochastic space. First, we proposed a formulation in order to compute the decomposition of high-order statistics. The idea is to compute the most influential parameters for high orders permitting to improve the sensitivity analysis. Second objective is to illustrate the correlation between the high-order functional decomposition and the PC-based techniques, thus displaying how to compute each term from a numerical point of view. Secondly, Basing on the Harten multiresolution framework in the stochastic space, we proposed a method allowing an adaptive refinement/derefinement in both physical and stochastic space for time dependent problems. As a consequence, an higher accuracy is obtained with a lower computational cost with respect to classical non-intrusive approaches, where the adaptivity is performed in the stochastic space only. Performances of this algorithm are tested on scalar Burgers equation and Euler system of equations, comparing with the classical Monte Carlo and Polynomial Chaos techniques.

Application of some of these techniques to tsunami simulations have been conducted.

6.9. Robust Design Optimization

Participant: Pietro Congedo [Corresponding member].

The Simplex-Simplex approach, that has been proposed in 2011, has been further developed. In particular, the algorithm has been improved yielding an evolved version of the Simplex2 approach, and the formulation has been extended to treat mixed aleatory/epistemic uncertainty. The resulting SSC/NM (Simplex Stochastic Collocation/Nelder-Mead) method, called Simplex2, is based on i) a coupled stopping criterion and ii) the use of an high-degree polynomial interpolation of the optimization space. Numerical results show that this method is very efficient for mono-objective optimization and minimizes the global number of deterministic evaluations to determine a robust design. This method is applied to some analytical test cases and a realistic problem of robust optimization of a multi-component airfoil. In this work, we present an extension of this method for treating epistemic uncertainty in the context of interval analysis approach. This method consists in a multi-scale strategy based on simplex space representation in order to minimize global cost of mixed epistemic-aleatory uncertainty quantification. This reduction is obtained i) by a coupled stopping criterion, ii) by an adaptive polynomial interpolation that could be used as a response surface in order to accelerate optimization convergence, iii) by a simultaneous min/max optimization sharing the same interpolating polynomials at each iteration [.....].

6.10. Multiphase flows

Participants: Rémi Abgrall [Corresponding member], Pietro Congedo, Maria-Giovanna Rodio, Harish Kumar.

We developed the numerical solver based on a DEM formulation modified for including viscous effects and a more complex equation of state for the vapor region. The method used is the DEM for the resolution of a reduced five equation model with the hypothesis of pressure and velocity equilibrium , without mass and heat transfer. This method results in a well-posed hyperbolic systems, allowing an explicit treatment of non conservative terms, without conservation error. The DEM method directly obtains a well-posed discrete equation system from the single-phase conservation laws, producing a numerical scheme which accurately computes fluxes for arbitrary number of phases. We considered two thermodynamic models , i.e. the SG EOS and the Peng-Robinson (PR) EOS. While SG allows preserving the hyperbolicity of the system also in spinodal zone, real-gas effects are taken into account by using the more complex PR equation. The higher robustness of the PR equation when coupled with CFD solvers with respect to more complex and potentially more accurate multi-parameter equations of state has been recently discussed. In this paper, no mass transfer effect is taken into account, thus the PR equation can be used only to describe the vapor behavior, while only the SG model is used for describing the liquid.

Another topic covered by Bacchus is about the numerical approximation of non conservative systems. One very interetsing example is obtained by the Kapila model, for which shock relations can be found from physical principles. Most, if not all, the know discretisation are at best stable, but do not converge under mesh refinement. We have proposed a way to do so by using some modifications of a Roe-like solver.

6.11. Parallel remeshing

Participants: Cécile Dobrzynski, Cédric Lachat, François Pellegrini [Corresponding member].

Our studies regarding parallel remeshing use a dedicated software framework called `PaMPA` (for "*Parallel Mesh Partitioning and Adaptation*"; see Section 5.6 for more details about it). This software, whose development started three years ago, allow one to describe distributed meshes in an abstract way.

The work carried out this year concerns the definition of suitable algorithms for performing remeshing in parallel, using a sequential remesher. To do so, areas suitable for remeshing (that is, cells for which a quality measurement routine indicates that remeshing is necessary) are grouped into boules of a size small enough to be handled by a sequential remesher, and big enough so that this remesher can do useful work on each of the boules. The core of the work is therefore to identify and build relevant boules, to send them to as many processors as possible, to remesh them sequentially, and to merge the remeshed boules into what remains of the original mesh. Then, areas that have not already been processed (e.g. areas at the interface of two or more boules) can be considered in turn, until all relevant cells have been considered. The structure and operations of `PaMPA` have been presented in [29].

Several algorithms have been experimented in order to build the boules. The one which proved the most efficient is based on a partitioning of an induced subgraph of the element graph, using the `PT-Scotch` tool which is already used for mesh redistribution. `PaMPA` has been interfaced with `MMG3D` in order to create a demonstrator for remeshing in parallel tetraedral meshes. A set of tetrahedral cube-shaped test meshes has been created, with a metric that coerces remeshing in the interior of the cubes. `PaMPA` was able to remesh a 12 million tetrahedral mesh into 18 million tetrahedra on 80 processors, yielding a quality equivalent to the one of the sequential remesher used alone. Scalability experiments on much larger test cases are in progress; yet, their quality will no longer be comparable to a sequential test case. This version of `PaMPA` will soon be released and made available to the community.

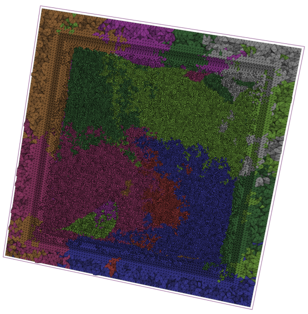


Figure 4. Cut of a 3D cube made of tetrahedra showing the effect of parallel remeshing by `PaMPA`.

6.12. Parallel graph remapping

Participants: Sébastien Fourestier, François Pellegrini [Corresponding member].

Last year, a set of new algorithms for sequential remapping and mapping with fixed vertices has been devised. These algorithms had been intergrated in the `Charm++` parallel environment, in the context of a collaboration with the Joint Laboratory for Petascale Computing (JLPC) between Inria and UIUC.

These algorithms have been integrated in version `6.0` of `Scotch`, which has been released in the beginning of December. This release also comprises new threaded formulations of the critical and most time-consuming algorithms used in graph partitioning, namely: graph coarsening and our diffusion-based method.

All the remapping algorithms that have been designed last year were meant to be easily parallelizable. The work of this year has been to derive and implement their parallel formulation. This is now the case, which completes this five year long work. These algorithms, which offer a quality similar to the one of the sequential algorithms, will be released in version `6.1` of `Scotch`.

6.13. Sparse matrix reordering for ILU solvers

Participants: Astrid Casadei, Sébastien Fourestier, François Pellegrini [Corresponding member].

In the context of ANR `PETALh`, our task is to find ways of reordering sparse matrices so as to improve the robustness of incomplete LU factorization techniques. The path we are following is to favor the diagonal dominance of the matrices corresponding to the subdomains of the Schur complement. Our studies aim at injecting some information regarding off-diagonal numerical values into nested dissection like reordering methods, so as to favor the preservation of high off-diagonal values into either the subdomains or the separators of Schur complement techniques.

This year, we have set-up a software testbed for experimenting such methods. It comprises a modified version of the `Scotch` sparse matrix ordering library for computing orderings and of the `HIPS` iterative sparse linear system solver for evaluating them. The text cases used are provided by the industrial partners of the `PETALh` project.

Our first experiments show that injecting information regarding off-diagonal terms can indeed improve convergence. However, many parameters have to be evaluated in a thorough experimentation plan. Since `Scotch` uses integer terms only, some scaling has to be performed, which imposes to determine how to scale the coefficients (type of scaling and range), whether to filter small values, etc. This work is in progress.

6.14. Subdomain decomposition

Participants: Astrid Casadei, François Pellegrini, Pierre Ramet [Corresponding member].

This work aims at finding subdomain decompositions that balance the sizes of off-diagonal contribution blocks.

In terms of graph partitioning, we have expressed this problem as a multi-constraint partitioning problem. In addition to bearing a weight that expresses the workload associated with its degrees of freedom, every graph vertex bears a second weight that holds the number of unknowns to which it is linked outside of its subdomain. Hence, in the nested dissection process, every time a separator is computed, this second weight is updated for each frontier vertex of the separated parts, before they are also recursively separated.

This year, we have set-up a software testbed for experimenting this approach. The `Scotch` sparse matrix ordering library has been modified so that graph vertices can bear multiple vertex weights. This required a slight change in the interfaces, but also modifications of the internal handling of graphs in many modules (nested dissection, graph coarsening, etc).

6.15. Development of a simulation code for rarefied gas flows

Participants: Luc Mieussens [Corresponding member], Florent Pruvost [IMB, engineer].

The simulation code `CORBIS` (rarefied gases in 2 space dimensions on structured meshes) has been entirely modified: modular form, use of the `git` version control system, modification to use unstructured meshes, MPI/OpenMP hybrid parallelization. Very good performance in terms of scalability and efficiency have been obtained, up to 700 cores.

6.16. Numerical methods for high altitude aerodynamics

Participants: Luc Mieussens [Corresponding member], N. Hérouard [CEA-CESTA, PhD].

In collaboration with CEA-CESTA, we have worked on the following subjects.

- A new method to generate locally refined velocity grids has been proposed. Very high performance improvement have been obtained (acceleration of the CPU time by a ratio around 30 for 3D computations). This work has been published in the proceedings of the 28th Symposium on rarefied Gas Dynamics, and is the subject of a paper submitted for publication.
- The second order Discontinuous Galerkin method has been studied for a one-dimensional problem of rarefied gases: we have shown that this method is clearly more accurate and faster than our finite volume method (which was used up to fourth order). This study will be developed in 2013 (numerical analysis and application to 2D problems).

6.17. Numerical methods for rarefied gas flows around moving obstacles

Participants: Luc Mieussens [Corresponding member], G. Dechristé [IMB, PhD].

We have presented one of the first numerical simulation of the Crookes radiometer. This phenomenon, due to the thermal creep flow, has been simulated with a Cartesian grid approach, with a cut-cell techniques that allow for an accurate treatment of solid boundaries. This work has been published in the proceedings of the 28th Symposium on rarefied Gas Dynamics.

6.18. Fast numerical methods for rarefied gases

Participants: Luc Mieussens [Corresponding member], Stéphane Brull [IMB], L. Forestier-Coste [IMB, Post Doc].

We have proposed a new method to discretize kinetic equations. It is basedd on a discretization of the velocity variable which is local in time and space. This induces an important gain in term of memory storage and CPU time, at least for 1D problems (this work has been rpesented in a paper submitted for publication). Two-dimensional extensions are under development.

6.19. Asymptotic Preserving schemes for the linear transport

Participant: Luc Mieussens [Corresponding member].

We have shown that the recent method “Unified Gas Kinetic Scheme”, proposed by K. Xu to simulated multiscale rarefied gas flows, can be extended to other fields, like radiative transfer. This approach, based on a simple finite volume technique, is very general and can be easily applied to complex geometries with unstructured meshes. This work has been presented in a paper submitted for publication.

BIPOP Project-Team

6. New Results

6.1. Multiple impacts modelling

Participants: Bernard Brogliato, Ngoc-Son Nguyen.

The work consists of studying two systems: the rocking block and tapered chains of balls, using the Darboux-Keller model of multiple impacts previously developed. The objectives are threefold: 1) show that the model predicts well the motion by careful comparisons with experimental data found in the literature, 2) study the system's dynamics and extract critical kinetic angles that allow the engineer to predict the system's gross motion, 3) develop numerical code inside the SICONOS platform that incorporates the model of multiple impact. Other works consist of analysing kinematic restitution laws based on the use of the kinetic energy metric. We also performed an analysis of the rocking block motion in terms of the kinetic angles between the two unilateral constraints. Results are in [21], [22] [55]. Another work is dedicated to analysing the influence of bilateral holonomic constraints on the well-posedness of the complementarity problem obtained from the (frictionless) unilateral constraints. Gauss' principle extension to this case is also analysed [20].

6.2. Discrete-time sliding mode control

Participants: Vincent Acary, Bernard Brogliato, Olivier Huber, Bin Wang.

This topic concerns the study of time-discretized sliding-mode controllers. Inspired by the discretization of nonsmooth mechanical systems, we propose implicit discretizations of discontinuous, set-valued controllers. This is shown to result in preservation of essential properties like simplcity of the parameters tuning, suppression of numerical chattering, reachability of the sliding surface after a finite number of steps, and disturbance attenuation by a factor h or h^2 [18]. In [23] we have provided a tutorial on similar types of systems like relay systems, and their relationships with other formalisms like complementarity systems, or switching dynamical systems. This follows in fact a research program proposed in [4].

6.3. Optimization

6.3.1. Optimization algorithms for large-scale machine learning problems, and applications in computer vision

Participant: Jérôme Malick.

This collaboration with Zaid Harchaoui (Inria, LEAR Team) has been growing since summer 2010. It also involves Miro Dudik (Microsoft Research NYC) and a student who just started his PhD in october 2012 (after his master with us).

The explosion of data that we are experiencing (Big Data) lead us to huge-scale learning problems, new challenges for statistical learning and numerical optimisation algorithms. For example, the new databases for image categorization are large-scale in the three dimensions (large number of exemples, high-dimension feature description, and large number of categories). The resulting learning problem is out of reach by standard optimization problems.

We developped a new approach exploiting the hidden underlying lower-dimension structure of this big data. We proposed a new family of algorithms (of the type coordinate results, or conditional gadient), whose iterations have an algorithmic complexity lower that an order compared to standard methods. For example, applied to learning problems with trace-norm penalization, our algorithm [26] exploit the atomic decomposition of the norm and compute only an approximate largest singular vector pair (instead of the whole singular value decomposition). Promising results [27] have been obtained on the image database Imagenet, where we show significant improvements compare to the state-of-the-art approaches (one-vs-rest strategies).

6.3.2. Semidefinite programming and combinatorial optimization

Participants: Nathan Krislock, Jérôme Malick.

We have worked with Frederic Roupin (Prof. at Paris XIII) on the use of semidefinite programming to solve combinatorial optimization problems to optimality. Within exact resolution schemes (branch-and-bound), “good” bounds are those with a “good” balance between tightness and computing times.

We proposed a new family of semidefinite bounds for 0-1 quadratic problems with linear or quadratic constraints [50]. The paradigm is to trade computing time for a (small) deterioration of the quality of the usual semidefinite bounds, in view of enhancing this efficiency in exact resolution schemes. Extensive numerical comparisons et tests showed the superior quality of our bounds, when embedded within branch-and-bound shemes, on standard test-problems (unconstrained 0-1 quadratic problems, heaviest k-subgraph problems, and graph bisection problems).

We have embedded the new bounds within branch-and-bound algorithms to solve 2 standard combinatorial optimization problems to optimality.

- *Max-cut.* We developed [34] an improved bounding procedure obtained by reducing two key parameters (the target level of accuracy and the stopping tolerance of the inner Quasi-Newton engine) to zero, and iteratively adding triangle inequality cuts. We also precisely analyzed its theoretical convergence properties. We show that our method outperform the state-of-the-art solver ([52]) on the large test-problems.
- *Heaviest k-subgraph problems*. Our previous work [51] takes advantage of the new bounds in their basic form to prune very well in the search tree. Its performances are then comparable with the best method (based on convex quadratic relaxation using CPLEX as an engine). Adapting and incorporating the tehniques we developped for the max-cut problem, we propose in [35] an big improveement of the first algorithm (up to 10 times faster). For the first time, we were able to solve exactly k-cluster instances of size 160. In practice, our method works particularly fine on the most difficult instances (with a large number of vertices, small density and small k).

Finally, we have worked on making our data sets available online together with a web interface for our solvers. We have also started working on a generic online semidefinite-based solver for binary quadratic problems using the generality of [50]. All this is publicly available on line at http://lipn.univ-paris13.fr/BiqCrunch/.

6.3.3. Unified theory of inaccurate bundle methods

Participants: Claude Lemaréchal, Welington Oliveira.

Convergence of bundle methods is an intricate subject. The situation is even worse in the inexact case, where many variants exist, each with its specific *ad hoc* proof techniques.

With C. Sagastizábal (Rio de Janeiro), we have developed a synthetic theory to single out the successive steps when proving convergence of a generic algorithm, as well as the specific hypotheses that they need. Our pattern covers all variants published so far and suggests a new one. The corresponding paper is being finalized.

6.3.4. Stabilizing marginal prices in electricity production

Participants: Claude Lemaréchal, Jérôme Malick, Sofia Zaourar.

Unit-commitment optimization problems in electricity production are large-scale, nonconvex and heterogeneous, but they are decomposable by Lagrangian duality. Realistic modeling of technical production constraints makes the dual objective function computed inexactly though. An inexact version of the bundle method has been dedicated to tackle this difficulty [48]. However, the computed optimal dual variables show a noisy and unstable behaviour, that could prevent their use as price indicator. We propose a simple and controllable way to stabilize the dual optimal solutions, by penalizing the total variation of the prices [36]. Our illustrations on the daily electricity production optimization of EDF show a strinking stabilization at a negligible cost.

6.4. Robotics

6.4.1. Hierarchic QP solver

Participants: Pierre-Brice Wieber, Dimitar Dimitrov.

We are working in collaboration with the LAAS-CNRS and the CEA-LIST on solving multi-objective Quadratic Programs with Lexicographic ordering: Hierarchic QPs [47]. The focus this year has been on enabling fast computations in the case of time-varying Hierarchic QPs through warm-starting the active set method. This has been possible by developing an active set method for lexicographic multi-objective ordering [44], [45]. The main difference with respect to classical active set methods is in the use of a "lexicographic" (sometimes called "multi-dimensional") Lagrange multiplier.

6.4.2. Modeling of human balance in public transports

Participants: Pierre-Brice Wieber, Zohaib Aftab.

Zohaib Aftab finished his PhD thesis in collaboration with IFSTTAR (previously INRETS) on modeling human balance in public transports. A Model Predictive Control scheme has been developed for the prediction of recovery motions, including ankle and hip strategies as well as stepping with adaptive step locations and timings [37]. This MPC scheme has been validated against a balance recovery scenario found in the biomechanics literature [38].

6.4.3. Model Predictive Control for Biped Walking

Participants: Pierre-Brice Wieber, Andrei Herdt, Jory Lafaye.

In collaboration with the DLR in Munich, we designed an MPC scheme for biped walking based on the "Capture Point". This is just a simple change of variable $\xi = x + \frac{1}{\omega}\dot{x}$ that transforms the second-order dynamics of the Center of Mass x of the robot into a cascade of two first-order dynamics, one stable and one unstable. This MPC scheme has been evaluated successfully on the DLR biped robot [49].

Since fast computations are always a key objective for feedback controllers, we designed a change of variable in the underlying QP in order to expose the specific structure between time-varying and time-invariant parts of the Hessian matrix and compute its Cholesky decomposition in an efficient way by pre-computing the decomposition of the time-invariant part.

6.5. Computational Toxicology

Participant:

It is now well recognized that toxicology has entered a new era. Previously mainly based on animal testing, toxicology is now turning to in vitro and in silico experiments. To assess the risk of chemicals but also to gather and to interpret the massive amounts of experimental data generated by modern toxicology, the development of mathematical and computational tools are essential. An important element in risk assessment of chemicals is the human bioaccumulative potential. We developed a predictive tool for human bioaccumulation assesment using a physiologically based toxicokinetic model [53].

6.6. Computational Biology

Participant:

Biological oscillations have attracted widespread interest from experimentalists, with the *in vivo* design of synthetic oscillators, and from mathematicians, with the study of limit cycles. Oscillations in protein concentrations or gene expression are supposed to be involved in the generation of the rhythms observed in the cell. In many situations, oscillations are originated by negative feedback loops. In [54] we have studied the oscillatory regimes of a negative feedback oscillator and derived the probability of having oscillations.

6.7. Mechanical rods

6.7.1. High-order models of mechanical rods

Participants: Florence Bertails-Descoubes, Romain Casati.

Reduced-coordinates models for rods such as the articulated rigid body model or the super-helix model [39] are able to capture the bending and twisting deformations of thin elastic rods while strictly and robustly avoiding stretching deformations. In this work we are exploring new reduced-coordinates models based on a higher-order geometry. Typically, elements are defined by a polynomial curvature function of the arc length, of degree $d \geq 1$. The main difficulty compared to the super-helix model (where $d = 0$) is that the kinematics has no longer a closed form. Last year we investigated the clothoidal case ($d = 1$) in the 2d case [19], relying on Romberg numerical integration. This year, in R. Casati's PhD's thesis, we extended this result to the full 3D case. The key idea was to integrate the rod's kinematics using power series expansion, and to design an accurate and efficient computational algorithm adapted to floating point arithmetics. Our method nicely propagates to the computation of the full dynamic of a linked chain of 3d clothoid. All these results will we submitted for publication early 2013.

6.7.2. Inverse modeling of mechanical rods

Participants: Florence Bertails-Descoubes, Alexandre Derouet-Jourdan.

Controlling the input shape of slender structures such as rods is desirable in many design applications (such as hairstyling, reverse engineering, etc.), but solving the corresponding inverse problem is not straightforward. In [43] we noted that reduced-coordinates models such as the super-helix are well-suited for static inversion in presence of gravity. The main difficulty then amounts to fitting a piecewise helix to an arbitrary input curve. This year in A. Derouet-Jourdan's PhD's thesis, we solved this problem by extending to 3d the floating tangents algorithm introduced in 2d in [43]. In this new method, only tangents are strictly interpolated while points are displaced in an optimal way so as to lie in a feasible configuration, *i.e.*, a configuration that is compatible with the interpolation by a helix. Our method proves to be efficient and robust as it can successfully handle large and complex datasets from real curve aquisitions, such as the capture of hair fibers or the magnetic field of a star. This result was submitted for publication to Computer-Aided Geometric Design in Spring, and is currently under minor revision.

6.8. High-accuracy time-stepping schemes

Participant: Vincent Acary.

To perform the numerical time integration of nonsmooth mechanical systems, the family of event-capturing time–stepping schemes are the most robust and efficient tools. Nevertheless, they suffer from several drawbacks : a) a low–order accuracy (at best at order one), b) a drift phenomena when the unilateral constraints are treated at the velocity level and c) a poor “energetic” behavior in terms of stabilizing the high–frequency dynamics. We fist proposed to improve the global order of accuracy over periods when the evolution is sufficiently smooth by mixing standard higher order schemes for Differential Algebraic equations and the Moreau–Jean's scheme [16]. We also proposed self-adapting schemes by applying time–discontinuous Galerkin methods to the measure differential equation in [24]. In order to satisfy in discrete time, the impact law and the constraints at the position and the velocity level, an adaptation of the well-known Gear–Gupta–Leimkuhler approach has been developed in [17]. Finally, the energetic behavior of the standard Moreau–Jean scheme has been addressed in [25] by developing a Newmark–type scheme for nonsmooth dynamics.

6.9. Dissipativity preserving methods

Participants: Vincent Acary, Bernard Brogliato.

This work concerns the analysis of so-called theta-methods applied to linear complementarity systems that are dissipative. Necessary and sufficient conditions for disspativity preservation after the time-discretization are derived (preservation of the stioarge function, the supply rate and the dissipation function). The possible state jumps are also analyzed [46]. It is shown that excepted when the system is state lossless and theta = 0.5, the conditions are very stringent.

6.9.1. Multivalued Lur'e dynamical systems

Participant: Bernard Brogliato.

Lur'e systems are quite popular in Automatic Control since the fifties. Set-valued Lur'e systems possess a static feedback nonlinearity that is a multivalued function. This study consists in the mathematical analysis (existence and uniqueness of solutions) and the stability analysis (Lyapunov stability, invariance principle) of classes of set-valued Lur'e systems, with applications in complementarity dynamical systems, relay systems, mechanical systems with dry friction, electrical circuits, etc.Our works in this field started in [40]. The results in [42] extend those in [41] with an accurate characterization of the maximal monotonicity of the central operator of these systems. Concrete and verifiable criteria are provided for the above classes (complementarity, relay systems).

CAD Team

6. New Results

6.1. Geometry Modeling and Processing

6.1.1. Relaxed lightweight assembly retrieval using vector space model

Participants: Kai-Mo Hu, Bin Wang, Jun-Hai Yong, Jean-Claude Paul.

Assembly searching technologies are important for the improvement of design reusability. However, existing methods require that assemblies possess high-level information, and thus cannot be applied in lightweight assemblies. In this paper, we propose a novel relaxed lightweight assembly retrieval approach based on a vector space model (VSM). By decomposing the assemblies represented in a watertight polygon mesh into bags of parts, and considering the queries as a vague specification of a set of parts, the resilient ranking strategy in VSM is successfully applied in the assembly retrieval. Furthermore, we take the scale-sensitive similarities between parts into the evaluation of matching values, and extend the original VSM to a relaxed matching framework. This framework allows users to input any fuzzy queries, is capable of measuring the results quantitatively, and performs well in retrieving assemblies with specified characteristics. To accelerate the online matching procedure, a typical parts based matching process, as well as a greedy strategy based matching algorithm is presented and integrated in the framework, which makes our system achieve interactive performance. We demonstrate the efficiency and effectiveness of our approach through various experiments on the prototype system. [19]

6.1.2. Calculating Jacobian coefficients of primitive constraints with respect to Euler parameters

Participants: Hai-Chuan Song, Jun-Hai Yong.

It is a fundamental problem to calculate Jacobian coefficients of constraint equations in assembly constraint solving because most approaches to solving an assembly constraint system will finally resort to a numerical iterative method that requires the first-order derivatives of the constraint equations. The most-used method of deriving the Jacobian coefficients is to use virtual rotation which is originally presented to derive the equations of motion of constrained mechanical systems. However, when Euler parameters are adopted as the state variables to represent the transformation matrix, using the virtual rotation will yield erroneous formulae of Jacobian coefficients. The reason is that Euler parameters are incompatible with virtual rotation. In this paper, correct formulae of Jacobian coefficients of geometric constraints with respect to Euler parameters are presented in both Cartesian coordinates and relative generalized coordinates. Experimental results show that our proposed formulae make Newton-Raphson iterative method converge faster and more stable. [22]

6.1.3. An extended schema and its production rule-based algorithms for assembly data exchange using IGES

Participants: Kai-Mo Hu, Bin Wang, Jun-Hai Yong.

Assembly data exchange and reuse play an important role in CAD and CAM in shortening the product development cycle. However, current CAD systems cannot transfer mating conditions via neutral file format, and their exported IGES files are heterogeneous. In this paper, a schema for the full data exchange of assemblies is presented based on IGES. We first design algorithms for the pre-and-post processors of parts based on solid model, in which the topologies are explicitly specified and will be referred by mating conditions, and then extend the IGES schema by introducing the Associativity Definition Entity and Associativity Instance Entity defined in IGES standard, so as to represent mating conditions. Finally, a production rule-based method is proposed to analyze and design the data exchange algorithms for assemblies. Within this schema, the heterogeneous representations of assemblies exported from different CAD systems can be processed appropriately, and the mating conditions can be properly exchanged. Experiments on the prototype system verify the robustness, correctness, and flexibility of our schema. [18]

6.1.4. Robust shape normalization of 3D articulated volumetric models

Participants: Yu-Shen Liu, Jun-Hai Yong, Jean-Claude Paul.

3D shape normalization is a common task in various computer graphics and pattern recognition applications. It aims to normalize different objects into a canonical coordinate frame with respect to rigid transformations containing translation, rotation and scaling in order to guarantee a unique representation. However, the conventional normalization approaches do not perform well when dealing with 3D articulated objects.

To address this issue, we introduce a new method for normalizing a 3D articulated object in the volumetric form. We use techniques from robust statistics to guide the classical normalization computation. The key idea is to estimate the initial normalization by using implicit shape representation, which produces a novel articulation insensitive weight function to reduce the influence of articulated deformation. We also propose and prove the articulation insensitivity of implicit shape representation. The final solution is found by means of iteratively reweighted least squares. Our method is robust to articulated deformation without any explicit shape decomposition. The experimental results and some applications are presented for demonstrating the effectiveness of our method. [24]

6.1.5. G^1 continuous approximate curves on NURBS surfaces

Participant: Jun-Hai Yong.

Curves on surfaces play an important role in computer aided geometric design. In this paper, we present a parabola approximation method based on the cubic reparameterization of rational Bézier surfaces, which generates G^1 continuous approximate curves lying completely on the surfaces by using iso-parameter curves of the reparameterized surfaces. The Hausdorff distance between the approximate curve and the exact curve is controlled under the user-specified tolerance. Examples are given to show the performance of our algorithm. [28]

6.1.6. The IFC-based path planning for 3D indoor spaces

Participant: Yu-Shen Liu.

Path planning is a fundamental problem, especially for various AEC applications, such as architectural design, indoor and outdoor navigation, and emergency evacuation. However, the conventional approaches mainly operate path planning on 2D drawings or building layouts by simply considering geometric information, while losing abundant semantic information of building components. To address this issue, this paper introduces a new method to cope with path planning for 3D indoor space through an IFC (Industry Foundation Classes) file as input. As a major data exchange standard for Building Information Modeling (BIM), the IFC standard is capable of restoring both geometric information and rich semantic information of building components to support lifecycle data sharing. The method consists of three main steps: (1) extracting both geometric and semantic information of building components defined within the IFC file, (2) discretizing and mapping the extracted information into a planar grid, (3) and finally finding the shortest path based on the mapping for path planning using Fast Marching Method. The paper aims to process different kinds of building components and their corresponding properties to obtain rich semantic information that can enhance applications of path planning. In addition, the IFC-based distributed data sharing and management is implemented for path planning. The paper also presents some experiments to demonstrate the accuracy, efficiency and adaptability of the method. Video demonstration is available from http://cgcad.thss.tsinghua.edu.cn/liuyushen/ifcpath/. [20]

6.1.7. Recovering Geometric Detail by Octree Normal Maps

Participants: Bin Wang, Jean-Claude Paul.

This paper presents a new approach for constructing normal maps that capture high-frequency geometric detail from dense models of arbitrary topology and are applied to the simplified version of the same models generated by any simplification method to mimic the same level of detail. A variant of loose octree scheme is used to optimally calculate the mesh normals. A B-spline surface fitting based method is employed to solve the issue of thin plate. A memory saving Breadth-First Search (BFS) order construction is designed. Furthermore, a speedup scheme that exploits access coherence is used to accelerate filtering operation. The proposed method can synthesize good quality images of models with extremely high number of polygons while using much less memory and render at much higher frame rate. [31]

6.1.8. An improved example-driven symbol recognition approach in engineering drawings

Participants: Hui Zhang, Ya-Mei Wen.

In this paper, an improved example-driven symbol recognition algorithm is proposed for CAD engineering drawings. First, in order to represent the structure of symbols more clearly and simply, we involve the text entity as one of the basic elements and redefine the relation representation mechanism, which is the foundation for the following algorithms. Then, the structure graph and a constrained tree can be established automatically for the target symbol, using the knowledge acquisition algorithm. In this process, the highest priority element is considered as the key feature, which will be regarded as the root node of the tree. The sequence of breadth first traveling will be recorded to be the recognition rule and saved in the symbol library. In the recognition process, the nodes with the same type as the key features can be located first in the drawing. Unnecessary matching calculations would be greatly reduced because of the accurate location. The other elements around, which satisfy the topology structure of the constrained tree, will be found next. The target symbol is recognized if all of the elements and constraints in the tree are found. Moreover, an extra preprocessing analysis approach is proposed to address repeat modes in a symbol. Thus, similar symbols can be recognized by one rule. We evaluate the proposed approach on the GREC databases and the real engineering drawings. The experimental results validate its effectiveness and efficiency. [17]

6.1.9. 3DMolNavi: A web-based retrieval and navigation tool for flexible molecular shape comparison

Participants: Yu-Shen Liu, Jean-Claude Paul.

Many molecules of interest are flexible and undergo significant shape deformation as part of their function, but most existing methods of molecular shape comparison treat them as rigid shapes, which may lead to incorrect measure of the shape similarity of flexible molecules. Currently, there still is a limited effort in retrieval and navigation for flexible molecular shape comparison, which would improve data retrieval by helping users locate the desirable molecule in a convenient way. To address this issue, we develop a web-based retrieval and navigation tool, named 3DMolNavi, for flexible molecular shape comparison. This tool is based on the histogram of Inner Distance Shape Signature (IDSS) for fast retrieving molecules that are similar to a query molecule, and uses dimensionality reduction to navigate the retrieved results in 2D and 3D spaces. We tested 3DMolNavi in the Database of Macromolecular Movements (MolMovDB) and CATH. Compared to other shape descriptors, it achieves good performance and retrieval results for different classes of flexible molecules. The advantages of 3DMolNavi, over other existing softwares, are to integrate retrieval for flexible molecular shape comparison and enhance navigation for user's interaction. [23]

6.1.10. Manifold-ranking based retrieval using k-regular nearest neighbor graph

Participants: Bin Wang, Kai-Mo Hu, Jean-Claude Paul.

Manifold-ranking is a powerful method in semi-supervised learning, and its performance heavily depends on the quality of the constructed graph. In this paper, we propose a novel graph structure named k-regular nearest neighbor (k-RNN) graph as well as its constructing algorithm, and apply the new graph structure in the framework of manifold-ranking based retrieval. We show that the manifold-ranking algorithm based on our proposed graph structure performs better than that of the existing graph structures such as k-nearest neighbor (k-NN) graph and connected graph in image retrieval, 2D data clustering as well as 3D model retrieval. In addition, the automatic sample reweighting and graph updating algorithms are presented for the relevance feedback of our algorithm. Experiments demonstrate that the proposed algorithm outperforms the state-of-the-art algorithms. [25]

6.2. Computer Graphics

6.2.1. Content-Based Color Transfer

Participants: Fuzhang Wu, Weiming Dong, Yan Kong, Xing Mei, Jean-Claude Paul, Xiaopeng Zhang.

This paper presents a novel content-based method for transferring the colour patterns between images. Unlike previous methods that rely on image colour statistics, our method puts an emphasis on high-level scene content analysis. We first automatically extract the foreground subject areas and background scene layout from the scene. The semantic correspondences of the regions between source and target images are established. In the second step, the source image is re-coloured in a novel optimization framework, which incorporates the extracted content information and the spatial distributions of the target colour styles. A new progressive transfer scheme is proposed to integrate the advantages of both global and local transfer algorithms, as well as avoid the over-segmentation artefact in the result. Experiments show that with a better understanding of the scene contents, our method well preserves the spatial layout, the colour distribution and the visual coherence in the transfer process. As an interesting extension, our method can also be used to re-colour video clips with spatially-varied colour effects. [26]

6.2.2. Large-scale forest rendering: Real-time, realistic, and progressive

Participants: Xiaopeng Zhang, Weiming Dong.

Real-time rendering of large-scale forest landscape scenes is important in many applications, such as video games, Internet graphics, and landscape and cityscape scene design and visualization. One challenge in the field of virtual reality is transferring a large-scale forest environment containing plant models with rich geometric detail through the network and rendering them in real time. We present a new framework for rendering large-scale forest scenes realistically and quickly that integrates extracting level of detail (LOD) tree models, rendering real-time shadows for large-scale forests, and transmitting forest data for network applications. We construct a series of LOD tree models to compress the overall complexity of the forest in view-dependent forest navigation. A new leaf phyllotaxy LOD modeling method is presented to match leaf models with textures, balancing the visual effect and model complexity. To progressively render the scene from coarse to fine, sequences of LOD models are transferred from simple to complex. The forest can be rendered after obtaining a simple model of each tree, allowing users to quickly see a sketch of the scene. To improve client performance, we also adopt a LOD strategy for shadow maps. Smoothing filters are implemented entirely on the graphics processing unit (GPU) to reduce the shadows' aliasing artifacts, which creates a soft shadowing effect. We also present a hardware instancing method to render more levels of LOD models, which overcomes the limitation of the latest GPU that emits primitives into only a limited number of separate vertex streams. Experiments show that large-scale forest scenes can be rendered with smooth shadows and in real time. [14]

6.2.3. Fast Multi-Operator Image Resizing and Evaluation

Participants: Weiming Dong, Xiaopeng Zhang, Jean-Claude Paul.

Current multi-operator image resizing methods succeed in generating impressive results by using image similarity measure to guide the resizing process. An optimal operation path is found in the resizing space. However, their slow resizing speed caused by inefficient computation strategy of the bidirectional patch matching becomes a drawback in practical use. In this paper, we present a novel method to address this problem. By combining seam carving with scaling and cropping, our method can realize content-aware image resizing very fast. We define cost functions combing image energy and dominant color descriptor for all the operators to evaluate the damage to both local image content and global visual effect. Therefore our algorithm can automatically find an optimal sequence of operations to resize the image by using dynamic programming or greedy algorithm. We also extend our algorithm to indirect image resizing which can protect the aspect ratio of the dominant object in an image. [16]

6.2.4. Easy modeling of realistic trees from freehand sketches

Participant: Xiaopeng Zhang.

Creating realistic 3D tree models in a convenient way is a challenge in game design and movie making due to diversification and occlusion of tree structures. Current sketch-based and image-based approaches for fast tree modeling have limitations in effect and speed, and they generally include complex parameter adjustment, which brings difficulties to novices. In this paper, we present a simple method for quickly generating various 3D tree models from freehand sketches without parameter adjustment. On two input images, the user draws strokes representing the main branches and crown silhouettes of a tree. The system automatically produces a 3D tree at high speed. First, two 2D skeletons are built from strokes, and a 3D tree structure resembling the input sketches is built by branch retrieval from the 2D skeletons. Small branches are generated within the sketched 2D crown silhouettes based on self-similarity and angle restriction. This system is demonstrated on a variety of examples. It maintains the main features of a tree: the main branch structure and crown shape, and can be used as a convenient tool for tree simulation and design. [21]

6.2.5. Real-time ink simulation using a grid-particle method

Participants: Shibiao Xu, Xing Mei, Weiming Dong, Xiaopeng Zhang.

This paper presents an effective method to simulate the ink diffusion process in real time that yields realistic visual effects. Our algorithm updates the dynamic ink volume using a hybrid grid-particle method: the fluid velocity field is calculated with a low-resolution grid structure, whereas the highly detailed ink effects are controlled and visualized with the particles. To facilitate user interaction and extend this method, we propose a particle-guided method that allows artists to design the overall states using the coarse-resolution particles and to preview the motion quickly. To treat coupling with solids and other fluids, we update the grid-particle representation with no-penetration boundary conditions and implicit interaction conditions. To treat moving "ink-emitting" objects, we introduce an extra drag-force model to enhance the particle motion effects; this force might not be physically accurate, but it proves effective for producing animations. We also propose an improved ink rendering method that uses particle sprites and motion blurring techniques. The simulation and the rendering processes are efficiently implemented on graphics hardware at interactive frame rates. Compared to traditional fluid simulation methods that treat water and ink as two mixable fluids, our method is simple but effective: it captures various ink effects, such as pinned boundaries and filament patterns, while still running in real time, it allows easy control of the animation, it includes basic solid-fluid interactions, and it can address multiple ink sources without complex interface tracking. Our method is attractive for animation production and art design.

6.2.6. Image zooming using directional cubic convolution interpolation

Participant: Weiming Dong.

Image-zooming is a technique of producing a high-resolution image from its low-resolution counterpart. It is also called image interpolation because it is usually implemented by interpolation. Keys' cubic convolution (CC) interpolation method has become a standard in the image interpolation field, but CC interpolates indiscriminately the missing pixels in the horizontal or vertical direction and typically incurs blurring, blocking, ringing or other artefacts. In this study, the authors propose a novel edge-directed CC interpolation scheme which can adapt to the varying edge structures of images. The authors also give an estimation method of the strong edge for a missing pixel location, which guides the interpolation for the missing pixel. The authors' method can preserve the sharp edges and details of images with notable suppression of the artefacts that usually occur with CC interpolation. The experiment results demonstrate that the authors'method outperforms significantly CC interpolation in terms of both subjective and objective measures. [30]

CAGIRE Team

6. New Results

6.1. Low Mach number flows simulations issue

The time-step dependency and the scaling of the pressure-velocity coupling suitable for unsteady calculations of low Mach number flows including acoustic features has been identified in the Momentum Interpolation approach. It has been shown that the proper form of the inertia term in the transporting velocity definition is related to the time-step independency of the steady state. The suitable scaling of the pressure gradient dissipation has been used to suggest a modification of AUSM+-up that allows acoustic simulations of low Mach number flows. The accuracy improvement when the solution is compared to the one of the original AUSM+-up scheme indicates that the scaling identified in the Momentum Interpolation approach can be applied with advantage to Godunov-type schemes [3].

6.2. Experimental results

The MAVERIC test facility has been significantly upgraded with the acquisition of a complete GPU-based system (hardware+software) that speeds up by a factor of 10 the processing of the PIV data. The strong sensitivity of the flow topology to the presence of an acoustic standing wave in the cross-flow has been clearly evidenced. The presently available measurements give already the possibility of extracting numerous velocity profiles for a future fruitful LES assessment. The dedicated 1-jet experiment for DNS assessment will start at the beginning of 2013 [8].

CALVI Project-Team

6. New Results

6.1. Mathematical analysis of kinetic models

6.1.1. Gyrokinetic and Finite Larmor radius approximations

Participants: Mihai Bostan, Céline Caldini, Emmanuel Frénod, Mathieu Lutz.

In a work in progress by E. Frénod and M. Lutz, the deduction of the Geometrical Gyro-Kinetic Approximation, which was originally obtained by Littlejohn in [75], [76], [77] using a physical approach which was mathematically formal, is done. The rigorous mathematical theory is built and explained in a form for providing it, especially, for analysts, applied mathematicians and computer scientists.

In the Note [16], we present the derivation of the finite Larmor radius approximation, when collisions are taken into account. We concentrate on the Boltzmann relaxation operator whose study reduces to the gyroaverage computation of velocity convolutions, which are detailed here. We emphasize that the resulting gyroaverage collision kernel is not local in space anymore and that the standard physical properties (i.e., mass balance, entropy inequality) hold true only globally in space and velocity. This work is a first step in this direction and it will allow us to handle more realistic collisional mechanisms, like the Fokker-Planck or Fokker-Planck-Landau kernels.

The subject matter of the paper [34] concerns the derivation of the finite Larmor radius approximation, when collisions are taken into account. Several studies are performed, corresponding to different collision kernels. The main motivation consists in computing the gyroaverage of the Fokker-Planck-Landau operator, which plays a major role in plasma physics. We show that the new collision operator enjoys the usual physical properties ; the averaged kernel balances the mass, momentum, kinetic energy and dissipates the entropy.

6.1.2. Singularities of the stationary Vlasov–Poisson system in a polygon

Participant: Simon Labrunie.

In collaboration with Fahd Karami (Université Cadi Ayyad, Morocco) and Bruno Pinçon (Université de Lorraine and project-team CORIDA), we conducted in [43] a theoretical and numerical study of the so-called “point effect” in plasma physics. The model (stationary Vlasov–Poisson system with external potential) corresponds a fully ionised plasma considered on a time scale much smaller than that of ions, but much larger than that of electrons. It appears as the relevant non-linear generalisation of the electrostatic Poisson equation. This may be a first step toward a quasi-equilibrium model valid on a larger time scale, where the equilibrium description of the electrons would be coupled to a kinetic or fluid model for the ions. This approximation is classical in plasma physics. We proved a general existence result for our model in a bounded domain $\Omega \subset \mathbb{R}^N$, which is not assumed to be smooth. When Ω is a polygonal domain of $\mathbb{R}^2$, we described the singular behavior of the solution near a reentrant corner. In the important case of the Maxwell–Boltzmann distribution, we established a link between various asymptotics of the problem and the (suitably extended) theory of large solutions to nonlinear elliptic problems (for a review of this theory, see e.g. [50]). This allowed us to determine the the dependence of the singularity coefficients on the parameters of the problem, such as the total mass of the distribution, or the boundary conditions of the potential. Numerical tests confirmed the theory.

6.2. Two-Scale Asymptotic-Preserving schemes

Participants: Nicolas Crouseilles, Emmanuel Frénod, Michaël Gutnic, Sever Hirstoaga.

In paper [20], we build a Two-Scale Macro-Micro decomposition of the Vlasov equation with a strong magnetic field. This consists in writing the solution of this equation as a sum of two oscillating functions with circumscribed oscillations. The first of these functions has a shape which is close to the shape of the Two-Scale limit of the solution and the second one is a correction built to offset this imposed shape. The aim of such a decomposition is to be the starting point for the construction of Two-Scale Asymptotic-Preserving schemes.

During CEMRACS 2011, we have started the project to test on a simplified model the Two-Scale Asymptotic-Preserving Schemes. The model, a two dimensional in phase space Vlasov-Poisson equation with small parameter, is used for a long time simulation of a beam in a focusing channel. This work was already done in [68] in the case where the solution is approximated by the two scale limit. The first goal is to improve this approximation, by going further, to the first order one; this was done in [41]. The second goal is to replace this approximation by an exact decomposition, using the macro-micro framework. This last approach will permit to treat the case of a not necessary small parameter. In order to accomplish the first task we have writen a Particle-In-Cell code in SeLaLib.

6.3. Development of numerical methods

Participants: Morgane Bergot, Anaïs Crestetto, Nicolas Crouseilles, Pierre Glanc, Michel Mehrenberger, Hocine Sellama, Eric Sonnendrücker, Christophe Steiner.

The work [19] is devoted to the numerical simulation of the Vlasov equation in the fluid limit using particles. To that purpose, we first perform a micro-macro decomposition as in [53] where asymptotic preserving schemes have been derived in the fluid limit. In [53], a uniform grid was used to approximate both the micro and the macro part of the full distribution function. Here, we modify this approach by using a particle approximation for the kinetic (micro) part, the fluid (macro) part being always discretized by standard finite volume schemes. There are many advantages in doing so: (i) the so-obtained scheme presents a much less level of noise compared to the standard particle method; (ii) the computational cost of the micro-macro model is reduced in the fluid regime since a small number of particles is needed for the micro part; (iii) the scheme is asymptotic preserving in the sense that it is consistent with the kinetic equation in the rarefied regime and it degenerates into a uniformly (with respect to the Knudsen number) consistent (and deterministic) approximation of the limiting equation in the fluid regime.

In [39] we present finite volumes schemes for the numerical approximation of the one-dimensional Vlasov-Poisson equation (FOV CEMRACS 2011 project). Stability analysis is performed for the linear advection and links with semi-Lagrangian schemes are made. Finally, numerical results enable to compare the different methods using classical plasma test cases.

In [40], we test an innovative numerical scheme for the simulation of the guiding-center model, of interest in the domain of plasma physics, namely for fusion devices. We propose a 1D Discontinuous Galerkin (DG) discretization, whose basis are the Lagrange polynomials interpolating the Gauss points inside each cell, coupled to a conservative semi-Lagrangian (SL) strategy. Then, we pass to the 2D setting by means of a second-order Strang splitting strategy. In order to solve the 2D Poisson equation on the DG discretization, we adapt the spectral strategy used for equally-spaced meshes to our Gauss-point-based basis. The 1D solver is validated on a standard benchmark for the nonlinear advection; then, the 2D solver is tested against the swirling deformation ow test case; finally, we pass to the simulation of the guiding-center model, and compare our numerical results to those given by the Backward Semi-Lagrangian method.

In [44] we have developed the guiding-center model in polar coordinates; numerical issues/difficulties can be tackled in such a test case which thus may be viewed as a first intermediate step between a classical center guide simulation in a 2D cartesian mesh and a 4D drift kinetic simulation.

In [25] and [28], we are interested in the numerical solution of the collisionless kinetic or gyrokinetic equations of Vlasov type needed for example for many problems in plasma physics. Different numerical methods are classically used, the most used is the Particle In Cell method, but Eulerian and Semi-Lagrangian (SL) methods that use a grid of phase space are also very interesting for some applications. Rather than using a uniform mesh of phase space which is mostly done, the structure of the solution, as a large variation of the gradients on different parts of phase space or a strong anisotropy of the solution, can sometimes be such that it is more interesting to use a more complex mesh. This is the case in particular for gyrokinetic simulations for magnetic fusion applications. We develop here a generalization of the Semi-Lagrangian method on mapped meshes. Classical Backward Semi-Lagrangian methods (BSL), Conservative Semi-Lagrangian methods based on one-dimensional splitting or Forward Semi-Lagrangian methods (FSL) have to be revisited in this case of mapped meshes. We consider here the problematic of conserving exactly some equilibrium of the distribution function, by using an adapted mapped mesh, which fits on the isolines of the Hamiltonian. This could be useful in particular for Tokamak simulations where instabilities around some equilibrium are investigated. We also consider the problem of mass conservation. In the cartesian framework, the FSL method automatically conserves the mass, as the advective and conservative form are shown to be equivalent. This does not remain true in the general curvilinear case. Numerical results are given on some gyrokinetic simulations performed with the GYSELA code and show the benefit of using a mass conservative scheme like the conservative version of the FSL scheme. Inaccurate description of the equilibrium can yield to spurious effects in gyrokinetic turbulence simulations. Also, the Vlasov solver and time integration schemes impact the conservation of physical quantities, especially in long-term simulations. Equilibrium and Vlasov solver have to be tuned in order to preserve constant states (equilibrium) and to provide good conservation property along time (mass to begin with). Several illustrative simple test cases are given to show typical spurious effects that one can observes for poor settings. We explain why Forward Semi-Lagrangian scheme bring us some benefits. Some toroidal and cylindrical GYSELA runs are shown that use FSL.

In [12] we present the Semi-Lagrangian method which is composed by essentially two ingredients : the computation of the characteristics along which the distribution function is constant and the interpolation step. We analyse high order schemes in time based on directional splitting, which are a succession of linear transport steps. We then study the Semi-Lagrangian methods in this particular case and we make the link between different formulations. We also obtain a convergence theorem for the Vlasov-Poisson system in this framework, which remains valid in the case of small displacements. We then develop this type of methods in a more general framework, by using one dimensionnal conservative splitting. We also consider a discontinuous Galerkin variant of such schemes. In a last part, we study the gyroaverage operator which appears in plasma physics by taking care of finite Larmor radius corrections. Finally, we discuss the problematic of zero discrete divergence which gives a compatibility between field computations and the numerical method of transport.

6.4. Finite Element Methods

6.4.1. Gyrokinetic quasi-neutrality equation

Participants: Nicolas Crouseilles, Eric Sonnendrücker.

In [21], a new discretization scheme of the gyrokinetic quasi-neutrality equation is proposed. We discretised the gyrokinetic Poisson equation using arbitrary order spline finite elements which enables to accommodate more complex domains. Moreover in standard polar coordinates we developed a fast solver which is comparable in computational time to the original FFT-second order finite differences, but can become more efficient for higher order as fewer grid points are needed for the same accuracy.

6.4.2. Dissipative boundary conditions for finite element codes

Participants: Philippe Helluy, Laurent Navoret, Eric Sonnendrücker.

We are developing finite-element codes for the Vlasov-Poisson system that would be able to capture the filamentation phenomenon. The filamentation is a mechanism that transfers the space fluctuations of the distribution function to high frequency oscillations in the velocity direction. For stability purpose, most numerical schemes contain dissipation that may affect the precision of the finest oscillations that could be resolved. In [60], [61], [62] Eliasson constructs a non reflecting and dissipative condition for the Fourier-transformed Vlasov-Poisson system. The condition enables the high velocity-frequency oscillations to leave the computational domain in a clean way.

We are currently developing a finite-element code based on this dissipative boundary condition. The code is part of the Selalib library. We also propose an approximation of the Eliasson method, based on the Béranger's PML formalism. Contrary to the original boudary conditions that requires a space Fourier transformation, this method is local and thus could be extended to higher dimensionnal problems and more complex geometries.

6.4.3. High order finite element methods for Maxwell

Participants: Stéphanie Salmon, Eric Sonnendrücker.

In paper [23], we study high order discretization methods for solving the Maxwell equations on hybrid triangle-quad meshes. We have developed high order finite edge element methods coupled with different high order time schemes and we compare results and efficiency for several schemes. We introduce in particular a class of simple high order low dissipation time schemes based on a modified Taylor expansion.

6.5. Waterbag models: analysis and simulations

Participant: Nicolas Besse.

In paper [33], we revisit the linear theory of kinetic flute-like modes such as ionic instabilities by using the exact geometric reduction of Vlasov equation yielded by waterbag invariants which are reminiscent to the geometric Liouville invariants. The waterbag representation of the statistical distribution function of particles can be viewed as a special class of exact weak solution of the Vlasov equation, allowing to reduce this latter into a set of hydrodynamic equations (with the complexity of a multi-fluid model) while keeping its kinetic features (Landau damping and resonant wave-particle interaction). For high toroidal-number-mode, from ballooning transformation and multi-scale WKB-type analysis, we demonstrate that one can construct eigenmode solutions of the two-dimensional integro-differential gyrowaterbag operator by solving a nested one-dimensional Fredholm-type integral equation. Qualitatively, the solution of the nested one-dimensional Fredholm-type equation is equivalent to first solving for the mode structure along the field lines locally in radius, and then constructing the two-dimensional global mode structure through a radially weighted superposition of local solutions. The radial weighted function is solution of a Schrödinger equation or a Riccati equation in the dual space. Solving the linear turning points problem and using connection formulas, the global dispersion relation arises from the WKB-type phase integral quantization condition which involves the local eigenfrequency. Finally we perform the spectral analysis of the nested one-dimensional Fredholm-type operator which constitutes a meromorphic family of compact operators and extend all the results proved for unstable eigenmodes to stable and damped ones by analytic continuation.

In paper [36], we present two new codes devoted to the study of ion temperature gradient (ITG) driven plasma turbulence in cylindrical geometry using a drift-kinetic multi-water-bag model for ion dynamics. Both codes were developed to complement the Runge-Kutta semi-lagrangian multi-water-bag code GMWB3D-SLC described in [55]. The CYLGYR code is an eigenvalue solver performing linear stability analysis from given mean radial profiles. It features three resolution schemes and three parallel velocity response models (fluid, multi-water-bag, continuous Maxwellian). The QUALIMUWABA quasi-linear code is an initial value code allowing the study of zonal flow influence on drift-waves dynamics. Cross-validation test performed between the three codes show good agreement on both temporal and spatial characteristics of unstable modes in the linear growth phase.

In paper [32], we first present the derivation of the anisotropic Lagrangian averaged gyrowaterbag continuum (LAGWBC) equations. The gyrowaterbag continuum can be viewed as a special class of exact weak solution of the Vlasov-gyrokinetic equation, allowing to reduce this latter into an infinite dimensional set of hydrodynamic equations (i.e. an infinite dimensional hyperbolic system of first-order conservation laws in several space dimensions with non-local fluxes) while keeping its kinetic features (Landau damping and nonlinear resonant wave-particle interaction). These models are very promising because they reveal to be very useful for analytical theory (such as the resolution of the eigenvalue problem for analytical description of various instabilities) and numerical simulations (when the continuum is converted into a small finite set of "fluid" or waterbag, the problem has the complexity of a multifluid model instead of a kinetic one) of laser-plasma and gyrokinetic physics (electrostatic turbulence problem). The gyrowaterbag waterbag continuum is derived from two phase-space variable reductions of the Vlasov equation through the existence of two underlying invariants. The first one, coming from physic properties of the dynamics (the fast gyromotion of particles around magnetic field lines) is adiabatic and called the magnetic moment. The second one, named "waterbag" and coming from geometric invariance property of the phase-space, is just the direct consequence of the Liouville Theorem and is reminiscent to the geometric Liouville invariant. In order to obtain the LAGWBC equations from the gyrowaterbag continuum we use an Eulerian variational principle and Lagrangian averaging techniques. Regarding to the original gyrowaterbag continuum, the LAGWBC equations show some additional properties and several advantages from the mathematical and physical viewpoints, which make this model a good candidate for describing accurately gyrokinetic turbulence in magnetically confined plasma. In the second part of this paper we prove local in time well-posedness of an approximated version of the anisotropic LAGWBC equations, that we call the “isotropic” LAGWBC equations, by using quasilinear PDE type methods and elliptic regularity estimates for several operators.

6.6. Simulations for Vlasov-Maxwell model

Participants: Anaïs Crestetto, Philippe Helluy.

In [37] (see also [11]), we present an implementation of a Vlasov-Maxwell solver for multicore processors. The Vlasov equation describes the evolution of charged particles in an electromagnetic field, solution of the Maxwell equations. We propose to solve the Vlasov equation by a Particle-In-Cell method (PIC), while the Maxwell system is computed by a Discontinuous Galerkin method. These methods are detailed, as well as the emission law for the particles and the implementation of the boundary conditions. We use the OpenCL framework, which allows our code to run on multicore processors or recent Graphic Processing Units (GPU). The key points of the implementation on this architecture are presented. We then study several numerical applications to two-dimensional test cases in cartesian geometry. The acceleration between the computation on a CPU and on a graphic card is very high, especially for the Maxwell part.

We have started a new software project called CLAC (for “Conservation Laws Approximation on many Cores”). This a 3D Discontinuous Galerkin solver, which runs on cluster of GPU’s, thanks to the OpenCL environment and the MPI library. CLAC is open source and developed in collaboration with the AxesSim company, a SME near Strasbourg. For the moment, it is applied to the Maxwell equations. But we plan to apply it to the MHD equations or mixed kinetic/fluid plasma models.

6.7. Free-streaming ELM formulae vs. Vlasov simulations

Participants: Sever Hirstoaga, Giovanni Manfredi.

One of the main challenges for future tokamak operation, such as ITER, is constituted by the large heat load on the divertor plates. The divertor surfaces are constantly bombarded with high-energy particles and may see their lifetime considerably reduced. The intensity of the particles and energy fluxes is particularly high during transient events known as edge-localised modes (ELMs). Our purpose here is to propose and investigate a kinetic model for ELMs.

The free-streaming model [69] is a simple analytical model for ELM transport in the scrape-off layer (SOL) of a tokamak. It is a force-free Vlasov equation with a source term for the ions distribution function (the Coulomb forces are ignored). Even though this model reproduces with good accuracy some of the main features of an ELM signal, it has two main drawbacks: (i) the self-consistent electric potential is not accounted for and (ii) only solutions for the ion distribution are considered.

In this contribution [24] we propose a set of modified free-streaming equations in order to overcome the above drawbacks. More precisely, some hypotheses on the Maxwellian initial condition lead to a model that includes the self-consistent electric potential. Assuming quasinetrality and using energy conservation we could derive analytical formulae for the electron quantities. This augmented free-streaming model was benchmarked to the Vlasov-Poisson simulations reported in [78]. The match is encouragingly good, thus justifying the applicability of the free-streaming approach.

Finally, from a computational point of view, transport in the SOL was studied by means of three different approaches – fluid, Vlasov and particle-in-cell (PIC). In spite of kinetic effects due to fast electrons which are not captured in the fluid code, the overall agreement between the codes was found to be quite satisfactory [22].

6.8. Full wave modeling of lower hybrid current drive in tokamaks

Participants: Pierre Bertrand, Takashi Hattori, Simon Labrunie, Jean Rodolphe Roche.

This work is performed in collaboration with Yves Peysson (DRFC, CEA Cadarrache). Since September 2012 this work is included in the ANR CHROME.

The aim of this project is to develop a finite element numerical method for the full-wave simulation of electromagnetic wave propagation in plasma. Full-wave calculations of the LH wave propagation is a challenging issue because of the short wave length with respect to the machine size. In the continuation of the works led in cylindrical geometry, a full toroidal description for an arbitrary poloidal cross-section of the plasma has been developed.

Since its wavelength λ at the LH frequency is very small as compared to the machine size R, a conventional full wave description represents a considerable numerical effort.Therefore, the problem is addressed by an appropriate mathematical finite element technique, which incorporates naturally parallel processing capabilities. It is based on a mixed augmented variational (weak) formulation taking account of the divergence constraint and essential boundary conditions, which provides an original and efficient scheme to describe in a global manner both propagation and absorption of electromagnetic waves in plasmas.

With such a description, usual limitations of the conventional ray tracing related to the approximation $\lambda << \phi_B << R$, where ϕ_B is the size of the beam transverse to the rf power flow direction, may be overcome. Since conditions are corresponding to $\lambda << \phi_B \sim R$, the code under development may be considered as a WKB full wave, dielectric properties being local.

This formulation provides a natural implementation for parallel processing, a particularly important aspect when simulations for plasmas of large size must be considered.

The domain considered is as near as possible of the cavity filled by a tokomak plasma. Toroidal coordinates are introduced. In our approach we consider Fourier decomposition in the angular coordinate to obtain stationary Maxwell equations in a cross-section of the tokamak cavity.

A finite element method is proposed for the simulation of time-harmonic electromagnetic waves in a plasma, which is an anisotropic medium. The approach chosen here is sometimes referred to as *full-wave modeling* in the literature: the original Maxwell's equations are used to obtain a second order equation for the time-harmonic electric field. These are written in a weak form using a augmented variational formulation (AVF), which takes into account the divergence. The variational formulation is then discretized using modified Taylor-Hood (nodal) elements.

During 2012 we have developed a domain decomposition method and a new behavior of the plasma density was considered in the code "FullWaveFEM". A analyze of the model considered, existence and unicity of solution, equivalence of the formulation for the domain decomposition formulation was completed in the frame of Takashi Hattori Phd thesis.

6.9. Nearby fields to plasma physics

6.9.1. Neutrino transport in supernova

Participant: Emmanuel Frénod.

In [31] we give an introduction to the Boltzmann equation for neutrino transport used in core collapse supernova models as well as a detailed mathematical description of the Isotropic Diusion Source Approximation (IDSA). Furthermore, we present a numerical treatment of a reduced Boltzmann model problem based on time splitting and finite volumes and revise the discretization of the IDSA for this problem. Discretization error studies carried out on the reduced Boltzmann model problem and on the IDSA show that the errors are of order one in both cases. By a numerical example, a detailed comparison of the reduced model and the IDSA is carried out and interpreted. For this example the IDSA modeling error with respect to the reduced Boltzmann model is numerically determined and localized.

In [30] we present Chapman–Enskog and Hilbert expansions applied to the O(v/c) Boltzmann equation for the radiative transfer of neutrinos in core collapse supernovae. Based on the Legendre expansion of the scattering kernel for the collision integral truncated after the second term, we derive the diffusion limit for the Boltzmann equation by truncation of Chapman–Enskog or Hilbert expansions with reaction and collision scaling. We also give asymptotically sharp results obtained by the use of an additional time scaling. The diffusion limit determines the diffusion source in the Isotropic Diffusion Source Approximation (IDSA) of Boltzmann's equation for which the free streaming limit and the reaction limit serve as limiters. Here, we derive the reaction limit as well as the free streaming limit by truncation of Chapman–Enskog or Hilbert expansions using reaction and collision scaling as well as time scaling, respectively. Finally, we motivate why limiters are a good choice for the definition of the source term in the IDSA.

6.9.2. Inverse problem governed by Maxwell equations

Participant: Jean Rodolphe Roche.

This work is performed in collaboration with José Herskovits Norman of UFRJ, Rio de Janeiro, Antonio André Novotny from the LNCC, Petropolis, both from Brazil and Alfredo Canelas from the University of the Republic, Montevideo, Uruguay.

The industrial technique of electromagnetic casting allows for contactless heating, shaping and controlling of chemical aggressive, hot melts. The main advantage over the conventional crucible shape forming is that the liquid metal does not come into contact with the crucible wall, so there is no danger of contamination. This is very important in the preparation of very pure specimens in metallurgical experiments, as even small traces of impurities, such as carbon and sulphur, can affect the physical properties of the sample. Industrial applications are, for example, electromagnetic shaping of aluminum ingots using soft-contact confinement of the liquid metal, electromagnetic shaping of components of aeronautical engines made of superalloy materials (Ni,Ti, ...), control of the structure solidification.

The electromagnetic casting is based on the repulsive forces that an electromagnetic field produces on the surface of a mass of liquid metal. In the presence of an induced electromagnetic field, the liquid metal changes its shape until an equilibrium relation between the electromagnetic pressure and the surface tension is satisfied. The direct problem in electromagnetic casting consists in determining the equilibrium shape of the liquid metal. In general, this problem can be solved either directly studying the equilibrium equation defined on the surface of the liquid metal, or minimizing an appropriate energy functional. The main advantage of this last method is that the resulting shapes are mechanically stable.

The inverse problem consists in determining the electric currents and the induced exterior field for which the liquid metal takes on a given desired shape. This is a very important problem that one needs to solve in order to define a process of electromagnetic liquid metal forming.

In a previous work we studied the inverse electromagnetic casting problem considering the case where the inductors are made of single solid-core wires with a negligible area of the cross-section. In a second paper we considered the more realistic case where each inductor is a set of bundled insulated strands. In both cases the number of inductors was fixed in advance, see [18]. In this year we aim to overcome this constraint, and look for configurations of inductors considering different topologies with the purpose of obtaining better results. In order to manage this new situation we introduce a new formulation for the inverse problem using a shape functional based on the Kohn-Vogelius criterion. A topology optimization procedure is defined by means of topological derivatives, a new method that simplifies computation issues was considered, see [35] and [29].

CASTOR Team

5. New Results

5.1. Simulations in plasma Physics

5.1.1. Fourier-spectral element approximation of a two fluid model of edge plasma

Participants: Richard Pasquetti, Sebastian Minjeaud.

We especially work on a two fluid physical model developed in close connection with Ph. Ghendrih (IRFM). It is based on the electrostatic assumption, i.e. the magnetic field is given (the magnetic field induced by the plasma itself is negligible), and on the hypothesis of electroneutrality (the density of ions and electrons are proportional). On the basis of the conservation equations of density, electron and ion velocities, electron and ion temperatures and electrical charges, a set of 10 non-linear coupled partial differential equations (PDE) can be set up. A high order Fourier-SEM (Spectral Element Method) code is currently developed. This Fourier-SEM code is now operational for the full set of PDEs in a 3D toroidal geometry. The torus section is discretized with quadrangular elements, within which the polynomial approximation degree is an input to the code. In time one uses an RK3 (third order Runge-Kutta) IMEX (Implicit-Explicit), so that the Lorentz terms are handled implicitly. The capability of this code to handle a strongly anisropic diffusion in a 3D toroidal geometry has already been tested. The Braginskii closure has been implemented. The Bohm boundary conditions at the plates are also considered. A parallel version of this code is currently developed. It remains to improve the robustness of our algorithms, i.e. to implement an efficient stabilization strategy. This could be based on the so-called spectral vanishing viscosity or entropy viscosity techniques. Up to our knowledge, this will be the first code that fully implements a two fluid ion-electron approximation (i.e. without using the drift velocity approximation), and the Braginskii closure of the governing equations.

5.1.2. Hydrodynamic model with strong Lorentz force

Participants: Audrey Bonnement, Hervé Guillard, Boniface Nkonga, Richard Pasquetti.

The thesis of A. Bonnement [1] was devoted to the development of a code based on the FluidBox/plaTo software of B. Nkonga and co-workers. It is based on a Finite volume / Finite element approach. This code is now operational in an axisymmetric geometry for a simplified PDE system in which the Lorentz force is approximated by a constant forcing field. Thus, the FluidBox/PlaTo code essentially solves the 3D axisymmetric Euler, Navier-Stokes or Braginskii PDEs to compute the ion density, momentum and energy. In the Braginskii system, the thermal diffusion and the kinematic viscosity are both non-linear and strongly anisotropic. A. Bonnement, who was co-directed by H. Guillard and R. Pasquetti, defended her thesis “Modélisation numérique bi-fluide du plasma de bord des tokamaks: application à ITER” in July 2012. A. Bonnement has provided a detailed description of the works carried out with the FluidBox/PlaTo code in her thesis manuscript. She has specially addressed one of the main difficulties related to simulations of tokamak plasmas, which is that the dynamic of the flows occurs in the vicinity of an equilibrium where the plasma pressure balances the Lorentz force. There are two ways to deal with this difficulty. The most common one in tokamak studies is to work with a set of governing equations such that this equilibrium is already contained in the formulation. This can be done by using formulations where the variables are indeed fluctuating departures from the equilibrium or by using special approximations as done in reduced MHD. The other way is purely numerical and consists to design a numerical method such that the equilibrium is an exact solution of the discrete equations. This has been the subject of the thesis of Audrey Bonnement in the framework of a finite volume method on non-structured meshes and where special Riemann solvers have been designed incorporating plasma equilibrium in the definition of the numerical fluxes. Combined with mesh refinement, this approach has been applied to some preliminary numerical experiments studying the effect of density perturbations (as a crude model of pellet injections) on the dynamics of the flow. At present, this approach is under evaluation to qualify its interest with respect to reduced MHD or formulations using a potential representation of the velocity field.

5.1.3. Finite volume methods in curvilinear system of coordinates

Participants: Hervé Guillard, Boniface Nkonga, Afeintou Sangam, Marco Bilanceri.

Finite volume methods are specialized techniques to approximate systems of conservation laws. The application of these methods to curvilinear systems of coordinate is however problematic because the space variation of the metric coefficients introduces artificial source terms. However it can be shown that whatever the curvilinear system used, a strong conservation form of the equations exists at the level of vector variables (but not at the level of the scalar components of the vectors in the curvilinear system due to the aforementioned space dependence of the metric coefficients). Based on this result, we have developed an original technique that uses an approximation of the vector form of the equation followed by local projection on the curvilinear system (here parallel to the poloidal magnetic field).

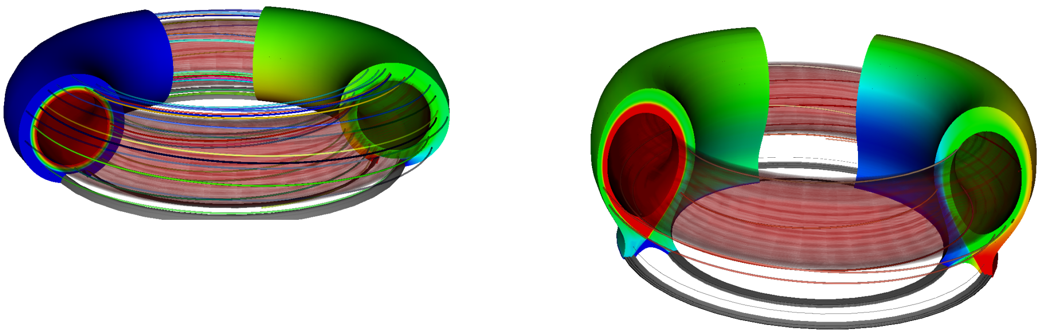


Figure 1. Density (left side) and parallel velocity (right side) color plots of the edge region of a tokamak with limiter (left plot) and tokamak with X point and divertor(right plot). Due to Bohm's boundary conditions, the parallel flux of out-flowing ions is supersonic on the limiter and divertor plates

This method has been applied to the approximation of a reduced MHD model using a decomposition of the velocity field into a parallel component and a perpendicular one given by the electric drift. The method is general and can be applied to any type of geometry. Figure 2 shows for instance the steady state density and parallel velocity fields in the edge region of a limiter tokamak (left) and of a divertor tokamak (right). Bohm's boundary conditions have been applied to the limiter and divertor plates producing a supersonic outflow velocity field.

5.1.4. Mesh singularities and triangular elements

Participants: Boniface Nkonga, Marie Martin, José Costa.

C1-finite elements as used for instance in the Jorek code are associated to isoparametric cubic-Bezier representation over quadrangles in the poloidal plane and sine-cosine Fourier expansion in the toroidal direction. Mesh singularities are associated to the structure imposed by the cubic-Bezier representation over quadrangles. In the context of the ANR-ANEMOS and in collaborations with IRFM and the Galaad team (Inria Sophia Antipolis), a geometrical toolbox is under development to manage these singularities and improve the alignment with equilibrium flux surfaces. As an alternative, we are also developing a more flexible C1-element over triangles using either Reduced-quintic (Bell) or quadratic Powell-Sabin polynomials. Optimal order of accuracy is achieved with simple boundary conditions. Many cycles of the "current hole" instability has been accurately reproduced. Additional improvement, with isoparametric formulation, will hopefully achieve this optimality for curved boundaries and improve mesh alignment to flux surfaces. We have investigated the possibility to use cubic splines representation in the toroidal direction. Indeed, for pellet injection the local resolution needed in the toroidal direction requires a large number of Fourier modes. This resolution need is very local, adapted splines representation can be more efficient. This solution is under analysis and structuration. First application is expected at end of 2013 with a possible update of Jorek in 2014.

5.1.5. Mesh adaptation Methods

Participants: Hubert Alcin [Projet Tropics], Alain Dervieux, Frédéric Alauzet [Projet Gamma, Inria-Rocquencourt].

This activity results from a cooperation between Gamma, Tropics, Castor, and Lemma company. See details in Tropics and Gamma activity reports. Its concerns Castor's subject through the current applications of mesh adaptation to flows with interfaces and to Large Eddy Simulation. It is also planned to use mesh adaptation for simplified plasma models in the context of ANEMOS ANR project.

5.1.6. Stabilization for finite / spectral element

Participants: Boniface Nkonga, Marie Martin, Richard Pasquetti, Sebastian Minjeaud.

Formulation of Reduced MHD eliminates fast acoustic waves but material, slow acoustic and Alfven waves are included in this model. On the other hand, finite element approximation, when applied to hyperbolic systems (with finite speed waves) needs additional control of the effect of unresolved scales. We have developed and validated a Taylor Galerkin Stabilizations of order 2 and 3 (TG2-TG3) for reduced MHD. This global approach has been implemented in a simplified form, validated and updated in the latest versions of the Jorek code. Even if significant improvements have been observed with this stabilization where only material and Alfven waves subscales are stabilized, more robustness is expected by taking into account slow acoustic waves. Stabilization techniques well adapted to high order approximations, like the spectral vanishing viscosity method or the entropy viscosity technique, remain to be implemented in the Fourier-SEM code.

5.1.7. Validity of the Reduced MHD and extensions

Participants: Hervé Guillard, Boniface Nkonga, Afeintou Sangam.

The available reduced MHD model in Jorek uses a set of assumption that can be reasonable close to the equilibrium and during the linear grow of instabilities. In order to obtain accurate and robust simulations of the nonlinear instabilities saturations, careful analysis and derivation of the reduced MHD has been performed, more mathematical derivations are under progress under the asymptotic analysis framework. It turns out that some of the neglected terms can be of relative importance for the saturation process when MHD instabilities move the plasma far from equilibrium.

5.1.8. High performance parallel computing

Participants: Hervé Guillard, Boniface Nkonga, Sebastian Minjeaud.

Applications under concern in this project needs to manage large meshes (10^7 to 10^9 nodes) and solve many huge sparse nonlinear systems. This makes the use of domain partitioning techniques unavoidable. In addition, since different numerical methodologies are under studies and evaluations in this project, we need to develop a quite general setting allowing the use of different data structures (element-oriented for FE vs edge-oriented for FV) and the possibility to consider different domain overlapping to efficiently communicate between processors. For this we develop the PaMPA (Parallel Mesh Partitioning and Adaptation) software in collaboration with the Bacchus team (Inria, Bordeaux). PaMPA is based on the PT-Scotch graph partitioning tool and allows on the fly mesh redistribution. Up to now, PaMPA has been tested on 10000 processors with a mesh of 20M tetrahedrons. Integration of PaMPA as an external library to the codes developed in this project is under progress and early results are promising. Similarly, the Fourier-SEM code is currently parallelized.

5.2. Optimisation and control for magnetic fusion plasmas

5.2.1. Evolutive equilibrium and transport coupling and optimization of scenarii

Participants: Jacques Blum, Cédric Boulbe, Afeintou Sangam, Gael Selig, Blaise Faugeras, Holger Heumann.

5.2.1.1. Research of optimal trajectories for the monitoring of Tokamak discharges

The direct equilibrium code CEDRES++ in its static version (resp.dynamic) computes for externally applied PF currents (resp. voltages) and given plasma current density profile the (resp. evolution of the) poloidal flux and the plasma free boundary. The research of optimal trajectories is the corresponding inverse problem : find externally applied currents (resp. voltages), such that the plasma reaches a certain desired state. This desired state is mainly (resp. the evolution of) a prescribed plasma boundary. We formulate these inverse problems as so-called optimal control problems, where the PF currents (voltages) are the so-called control variables and the poloidal flux the so-called state variable. Optimal control problems are optimization problems with PDE (partial differential equations) constraints. In our case, the Grad-Shafranov equation is the constraint and the functional to be minimized is a cost-function that measures the mismatch between the computed plasma boundary and the desired plasma boundary. The Sequential Quadratic Programming (SQP) method is known to be a very efficient algorithm for solving non-linear constrained optimization problems. We implemented in CEDRES++ the SQP method for the two cases of finding either currents or voltages that corresponds to a desired boundary or a desired evolution of the boundary. These implementations are built on the orignal Newton methods for the direct non-linear problems. For optimization problems it is of great importance that the Newton methods are 'real' Newton methods in the sense that the Newton matrices are real derivatives. In the original implementation of CEDRES++ these matrices were the discretization of analytic derivatives of the non-linear operators, hence not derivatives of the discrete problem. We had to rewrite large parts of the code to eliminate this problem. Further, we added an interface to the linear solver library UMFPACK. For the current mesh resolution level, the performance of this linear solver for the stationary problems, both in the direct and in the inverse versions, is superior to iterative linear solvers. In the case of the inverse non-stationary problem, the problem of finding voltages that correspond to a desired evolution, the memory requirements forbid the use of UMFPACK. There, we used Conjugate Gradient-type iterations. In the future, we will have to investigate if other types of iterative solvers are suitable and allow a certain parallelism that will speed up the simulation time.

5.2.1.2. A new method of coupling equilibrium and resistive diffusion equations

In the framework of Gael Selig's PhD thesis, the resistive diffusion equation has been incorporated in the evolutive equilibrium system of CEDRES++. This equation has as unknown variable the derivative of the poloidal flux with respect to the averaged minor radius of the magnetic surface. This choice was made instead of the poloidal flux itself because this is the quantity directly involved in the averaged Grad-Shafranov equation used to compute the FF' term and thus this allows us not to perform a supplementary numerical differentiation which might introduce some numerical instability. An algorithm based on a successive prediction and correction method is proposed in order to ensure the consistency between the evolution of the 2D poloidal flux in the equilibrium equation and the evolution of the poloidal flux in the 1D resistive diffusion equation. The algorithm guarantees that at the end of each time step the total plasma current Ip and the mean radius of the plasma have the same values in both systems (see fig.2). The convergence of this new code (called CEDRES-DIF) has been numerically validated and the method has been successfully compared by G. Selig to the CEDRES-CRONOS coupled code which uses another coupling algorithm.

5.2.1.3. Introduction of halo currents in the equilibrium resolution

When VDE (Vertical Displacement events) instabilities occur in a Tokamak, currents flow from the plasma to the machine vessel structures, and then return to the plasma. These currents are called halo currents . In turn, these currents induce forces on the wall when crossing with Tokamak poloidal and toroidal magnetic fields. Moreover, when VDE instabilities take place, the plasma hits the wall with all its energy. Therefore, it is worth understanding the contribution of halo currents to total plasma current and other related plasma parameters, particularly the distribution, magnitude, and temporal evolution of halo currents for large scale machine such as ITER. Even if halo currents are actually 3D phenomena, it is important to take into account their effects in 2D free boundary equilibrium codes. In halo region, the pressure can be considered as negligeable so that the current follows the magnetic field lines. The magnetic field satisfies the force free equation $jxB = 0, \nabla.B = 0$ which can be rewritten

$$-\Delta^* \Psi = \frac{1}{\mu_0 R} \frac{\partial}{\partial \Psi} f_H^2(\Psi)$$

in an axisymmetric configuration. The function $f_H(\Psi)$ is supposed to be known. This simple model has been implemented in CEDRES++ and first tests have been done. This first simplified model has to be improved to get more realistic simulations and to be validated. The choice of the function f_H, the value of the total halo current, the geometry and the size of the halo region need to be enhanced with respect to experimental data.

5.2.2. Equilibrium reconstruction and current density profile identification

Participants: Jacques Blum, Cédric Boulbe, Blaise Faugeras.

EQUINOX is a real-time equilibrium reconstruction code. It solves the equation satisfied by the poloidal flux in a computation domain, which can be the vacuum vessel for example, using a P1 finite element method and solves the inverse problem of the identification of the current density profile by minimizing a least square cost-function. It uses as minimal input the knowledge of the flux and its normal derivative on the boundary of the computation domain. It can also use supplementary constraints to solve the inverse problem: interferometric, polarimetric and MSE measurements. Part of the work reported here has been done in the frame of a RTM-JET contract.

5.2.2.1. Direct use of the magnetic measurements

Equinox was not originally designed to take as magnetic inputs directly the magnetic measurements, as it should be the case in the ITM, but some outputs from the real-time codes Apolo at ToreSupra and Xloc at JET. These codes provide Equinox with the values of the flux and its normal derivative on a closed contour defining the boundary of the computation domain (this contour can be the limiter for example). As a consequence the main difficulty arising in the objective of integrating the code Equinox in the ITM structure was to interpolate between the magnetic measurements (flux loops and poloidal B-probes) with a machine independent method. This has already been achieved by using toroidal harmonic functions, as a basis for the decomposition of the poloidal flux in the vacuum region, in complement to the contribution of the PF coils. This method can provide an alternative tool, comparable to APOLO (for Tore Supra) and FELIX (for JET), to compute the plasma boundary in real time from the magnetic measurements. Some twin experiments for WEST (Tore Supra upgrade) have been successfully conducted. In a first step the equivalents of magnetic measurements were generated using the FBE code CEDRES++. In a second step these measurements were used by the toroidal harmonics algorithm to reconstruct the plasma boundary. The results are very promising and the work on this subject is ongoing for JET.

5.2.2.2. Boundary conditions for EQUINOX

In the present version of EQUINOX the boundary condition is a flux condition (Dirichlet boundary condition) and the tangential component of the poloidal field is incorporated in the cost-function to be minimized. This is a constant criticism which is made on EQUINOX. The idea was to inverse these two boundary conditions in order to determine if this choice is determinant in the results. We tried to use the tangential poloidal field (Neumann boundary condition for the flux) as boundary condition for the boundary value problem, and to put the flux (or its tangential derivative linked to the normal component of the poloidal field) in the cost function. However no convincing results could be obtained because the numerical resolution of the boundary value problem associated with Neumann boundary conditions proved to be unstable. This might be explained by the fact that a compatibility condition has to be satisfied between the Neumann conditions and the current density in the plasma which evolves during the mixed fixed-point and optimization iterations.

5.2.2.3. Induced currents in EQUINOX

In a disruption when the total plasma current disappears, there are very important induced currents, for example in the toroidal pumped limiter. These currents are in the domain of resolution of EQUINOX. Therefore it is necessary to take them into account in the resolution of the equilibrium reconstruction problem. This has been tested on a Tore Supra disruption case. The mesh generation has been modified in order to incorporate the real structure of the limiter. The structure of the equations being solved in the code also had to be modified in order to take into account the measured induced currents.

5.3. Turbulence models

5.3.1. Hybrid RANS/LES models

Participants: Hubert Alcin [Tropics], Alain Dervieux, Bruno Koobus [University of Montpellier 2], Carine Moussaed [University of Montpellier 2], Maria-Vittoria Salvetti [University of Pisa], Stephen Wornom [Lemma].

The purpose of our works in hybrid RANS/LES is to develop new approaches for industrial applications of LES-based analyses. In the foreseen applications (aeronautics, hydraulics), the Reynolds number can be as high as several tenth millions, a far too large number for pure LES models. However, certain regions in the flow can be much better predicted with LES than with usual statistical RANS (Reynolds averaged Navier-Stokes) models. These are mainly vortical separated regions as assumed in one of the most popular hybrid model, the hybrid Detached Eddy Simulation model. Here, “hybrid” means that a blending is applied between LES and RANS. The french-italian team has designed a novel type of hybrid model. This year, a novel dynamic formulation has been introduced in our models and tested. the new model has been adapted to very high Reynolds number. Carine Moussaed has presented her results in ECCOMAS (Vienna). In our set of benchmark test cases which are also ECINADS test cases, the flow past a circular cylinder at Reynolds number from 3900 to 1 Million could be passed with improved predictions of main properties like mean drag, root mean square of lift fluctuation, and base pressure.

5.3.2. Acoustics

Participants: ILya Abalakin [IMM-Moscou], Alain Dervieux [Tropics], Alexandre Carabias [Tropics], Tatyana Kozubskaya [IMM-Moscow], Bruno Koobus [University of Montpellier 2].

A method for the simulation of aeroacoustics on the basis of hybrid RANS/LES models has been designed and developed by a cooperation between the Computational Aeroacoustics Laboratory (CAL) of Intitute for Mathematical Modeling at Moscow and Inria. Further applications has been developed by the Russian team from the two common numerical scheme, the Mixed-Element-Volume at sixth-order, and the quadratic reconstruction scheme. This year the cooperation is concentrated on the study by Alexandre Carabias of a new quadratic reconstruction scheme, which extends the one developed by Hilde Ouvrard and Ilya Abalakin. This year, this scheme is also introduced in the Gamma-Sciport mesh adaptation loop.

5.4. Environmental flows

Participants: Hervé Guillard, Boniface Nkonga, Marco Bilanceri, Maria-Vittoria Salvetti [University of Pisa, Italy], Imad Elmahi [University of Oudja, Morocco].

Mobile bed and sediment transport

The numerical approximation of a model coupling the shallow-water equations with a sediment transport equation for the morphodynamics have been studied. In shallow-water problems, time advancing can be carried out by explicit schemes. However, if the interaction with the mobile bed is weak, the characteristic time scales of the flow and of the sediment transport can be very different introducing time stiffness in the global problem. For this case, it is of great interest to use implicit schemes. The time integration stategy that we have devised is based on a defect-correction approach and on a time linearization, in which the flux Jacobians are computed through automatic differentiation. The aim of the present work is to investigate the behaviour of this time scheme in different situations related to environmental flows. This work has been published in [14] and is now applied to the study of the Nador Lagoon in Morocco.

CLASSIC Project-Team

5. New Results

5.1. Contributions earlier to 2012 but only published in 2012

Participants: Gérard Biau, Vincent Rivoirard, Gilles Stoltz, Olivier Catoni.

We do not discuss here the contributions provided by [16], [17], [11], [13], [14], [15], [18], since they were achieved in 2011 or earlier (but only published this year due to the reviewing and publishing process). Also, the book [25] (whose first edition was published in 2009) was augmented and revised for its second edition, published this year.

5.2. Extended journal versions written in 2012 of conference papers published in 2011

Participants: Sébastien Gerchinovitz, Gilles Stoltz.

We wrote extended journal papers of some conference papers discussed in previous annual activity reports; they correspond to refences [32], [19], [20].

5.3. Bayesian methods

Participants: Gérard Biau, Vincent Rivoirard.

5.3.1. The ABC method

Approximate Bayesian Computation (ABC for short) is a family of computational techniques which offer an almost automated solution in situations where evaluation of the posterior likelihood is computationally prohibitive, or whenever suitable likelihoods are not available. In the paper [29] Gérard Biau and his coauthors analyze the procedure from the point of view of k-nearest neighbor theory and explore the statistical properties of its outputs. They discuss in particular some asymptotic features of the genuine conditional density estimate associated with ABC, which is a new interesting hybrid between a k-nearest neighbor and a kernel method.

5.3.2. Semi-parametric version of the Bernstein-von Mises theorem

In [22], Vincent Rivoirard and Judith Rousseau study the asymptotic posterior distribution of linear functionals of the density by deriving general conditions to obtain a semi-parametric version of the Bernstein-von Mises theorem. The special case of the cumulative distributive function evaluated at a specific point is widely considered. In particular, they show that for infinite dimensional exponential families, under quite general assumptions, the asymptotic posterior distribution of the functional can be either Gaussian or a mixture of Gaussian distributions with different centering points. This illustrates the positive but also the negative phenomena that can occur for the study of Bernstein-von Mises results. In [22] Vincent Rivoirard and Judith Rousseau use convergence rates on Besov spaces established in [23].

5.4. Sequential learning

Participants: Pierre Gaillard, Gilles Stoltz.

5.4.1. Bandit problems

The article [30] revisits asymptotically optimal results of Lai and Robbins, Burnetas and Katehakis in a non-asymptotic way. A preliminary attempt was mentioned in the 2011 annual report; it was concerned (essentially) with the case of Bernoulli distributions over the arms. We achieve here the stated optimality of the regret bounds for larger models: regular exponential families; finitely supported distributions.

5.4.2. Theoretical results for the prediction of arbitrary sequences

We generalize and unify in [24] several notions of regret under a same banner: these include adaptive regret (regret against a fixed convex combination on subintervals of the time); shifting regret (regret against a slowly evolving target sequence of convex combinations); and discounted regret (when the instances are weighted with weights depending on how recent the instances are). We recover and sometimes improve some earlier bounds.

5.4.3. Forecasting of the production data of oil reservoirs

We applied our sequential aggregation techniques to a new data set, with IFP Energies nouvelles as a partner. The goal was to aggregate in a sequential fashion the forecasts made by some (about 100) base experts in order to predict some behaviors (gas/oil ratio, cumulative oil extracted, water cut) of the exploitation of some oil wells. Results were obtained with the help of an intern, Charles-Pierre Astolfi, and are described in the technical report [27] (to be transformed into a regular journal / conference paper next year).

5.5. Regression, classification, regression methods

Participants: Gérard Biau, Olivier Catoni, Ilaria Giulini.

5.5.1. Metric-based decision procedures

We know now that a good part of the statistical performance of regression and classification algorithms relies on the metric chosen to represent the proximity between the data points. Throughout his work, Gérard Biau became convinced that, well beyond the traditional distances, (dis)similarities and other self-reproducing kernel metrics, it is now necessary to attempt to define proximities generated by the sample itself. These metrics are inevitably random and probabilistic, and force us to rethink the nature of the estimates, as shown for example in the preliminary article [12].

5.5.2. Unsupervised classification in reproducing kernel Hilbert spaces

In her PhD started in September 2012, Ilaria Giulini uses dimension free estimates of the principal components of an i.i.d. sample of points in a Reproducing Kernel Hilbert Space to derive new unsupervised clustering algorithms based on the idea of dimension reduction by nonlinear coordinate smoothing along aggregated principal components. The dimension free estimates are obtained using PAC-Bayes bounds derived from thresholded exponential moments.

5.6. Sparsity and ℓ_1–regularization

Participant: Vincent Rivoirard.

5.6.1. For multivariate Hawkes processes

Motivated by statistical problems in neuroscience, Vincent Rivoirard and his coauthors study in [31] non-parametric inference for multivariate Hawkes processes depending on an unknown function to be estimated by linear combinations of a fixed dictionary. To select coefficients, they propose a Lasso-type methodology where data-driven weights of the penalty are derived from new Bernstein-type inequalities for martingales. Oracle inequalities are established under assumptions on the Gram matrix of the dictionary. Non-asymptotic probabilistic results are proven, which allows them to check these assumptions by considering general dictionaries based on histograms, Fourier or wavelet bases. They finally carry out a simulation study and compare their methodology with the adaptive Lasso procedure proposed by Zou. They observe an excellent behavior of their procedure with respect to the problem of supports recovery. Unlike adaptive Lasso of Zou, their tuning procedure is proven to be robust with respect to all the parameters of the problem, revealing its potential for concrete purposes in neuroscience, but also in other fields.

5.6.2. In the spherical convolution model

In [21], Thanh Mai Pham Ngoc and Vincent Rivoirard consider the problem of estimating a density of probability from indirect data in the spherical convolution model. They aim at building an estimate of the unknown density as a linear combination of functions of an overcomplete dictionary. The procedure is devised through a well-calibrated ℓ_1–penalized criterion. The dictionary approach allows to combine various bases and thus enhances estimates sparsity. They provide an oracle inequality under global coherence assumptions. Moreover, the calibrated procedure that they put forward gives very satisfactory results in the numerical study when compared with other procedures.

5.6.3. For semiparametric nonlinear mixed-effects models

Semiparametric nonlinear mixed-effects models (SNMMs) have been proposed as an extension of nonlinear mixed-effects models (NLMMs). These models are a good compromise and retain nice features of both parametric and nonparametric models resulting in more flexible models than standard parametric NLMMs. In [28], Vincent Rivoirard and his coauthors propose new estimation strategies in SNMMs. They propose a Lasso-type method to estimate the unknown nonlinear function. They derive oracle inequalities for this nonparametric estimator. They combine the two approaches in a general estimation procedure that they illustrate with simulations and through the analysis of a real data set of price evolution in on-line auctions.

5.7. Computational linguistics

Participants: Olivier Catoni, Thomas Mainguy.

In a forthcoming paper, Olivier Catoni and Thomas Mainguy study a new statistical model to learn the syntactic structure of natural languages from a training set made of written sentences. This model learns a new type of stochastic grammar and defines a statistical model on sentences. Global constraints are enforced, that set the approach apart from the family of Markov models. On the other hand, the grammar model generates outputs through a split and merge stochastic process that is more elaborate than the production rules defining a context free grammar. Experiments made on small corpora are very encouraging. Working on large corpora will require to speed up the algorithms used to implement the model as well as some code optimization.

COFFEE Project-Team (section vide)

COMMANDS Project-Team

6. New Results

6.1. Optimal control of partial differential equations

6.1.1. Optimal control of a semilinear parabolic equation with singular arcs

Participant: Frédéric Bonnans.

This paper, published as Inria report 8099 [25], develops a theory of singular arc, and the corresponding second order necessary and sufficient conditions, for the optimal control of a semilinear parabolic equation with scalar control applied on the r.h.s. We obtain in particular an extension of Kelley's condition, and the characterization of a quadratic growth property for a weak norm.

6.2. Trajectory optimization

6.2.1. First and second order optimality conditions for optimal control problems of state constrained integral equations

Participants: Frédéric Bonnans, Xavier Dupuis.

In this work performed with Constanza De La Vega (U. Buenos Aires), and published as Inria report 7961 [26], we deal with optimal control problems of integral equations, with initial-final and running state constraints. The order of a running state constraint is defined in the setting of integral dynamics, and we work here with constraints of arbitrary high orders. First and second-order necessary conditions of optimality are obtained, as well as second-order sufficient conditions.

6.2.2. Sensitivity analysis for relaxed optimal control problems with final-state constraints

Participants: Frédéric Bonnans, Laurent Pfeiffer.

In this work, performed with Oana Serea (U. Perpignan), and published as Inria report 7977 [27], we compute a second-order expansion of the value function of a family of relaxed optimal control problems with final-state constraints, parameterized by a perturbation variable. The sensitivity analysis is performed for controls that we call R-strong solutions. They are optimal solutions with respect to the set of feasible controls with a uniform norm smaller than a given R and having an associated trajectory in a small neighborhood for the uniform norm. In this framework, relaxation enables us to consider a wide class of perturbations and therefore to derive sharp estimates of the value function.

6.2.3. Sensitivity analysis for the outages of nuclear power plants

Participants: Frédéric Bonnans, Laurent Pfeiffer.

In this work, performed with Kengy Barty (EDF), and published as Inria report 7884 [24]. Nuclear power plants must be regularly shut down in order to perform refueling and maintenance operations. The scheduling of the outages is the first problem to be solved in electricity production management. It is a hard combinatorial problem for which an exact solving is impossible.

Our approach consists in modelling the problem by a two-level problem. First, we fix a feasible schedule of the dates of the outages. Then, we solve a low-level problem of optimization of elecricity production, by respecting the initial planning. In our model, the low-level problem is a deterministic convex optimal control problem.

Given the set of solutions and Lagrange multipliers of the low-level problem, we can perform a sensitivity analysis with respect to dates of the outages. The approximation of the value function which is obtained could be used for the optimization of the schedule with a local search algorithm.

6.2.4. Optimization of the anaerobic digestion of microalgae in a coupled process

Participant: Pierre Martinon.

In this work in collaboration with Terence Bayen (U. Monptellier) and Francis Mairet (Inria Sophia), submitted to ECC13 [30], we study the maximization of the production of methane in a bioreactor coupling an anaerobic digester and a culture of micro-algae limited by light. The decision parameter is the dilution rate which is chosen as a control, and we enforce periodic constraints in order to repeat the same operation every day. The system is gathered into a three-dimensional system taking into account a day-night model of the light in the culture of micro-algae. Applying Pontryagin maximum principle, the necessary conditions on optimal trajectories indicate that the control consists of bang and/or singular arcs. We provide numerical simulations by both direct and indirect methods, which show the link between the light model and the structure of optimal solutions.

6.3. Stochastic programming

6.3.1. Solving multi-stage stochastic mixed integer linear programs by the dual dynamic programming approach

Participants: Frédéric Bonnans, Zhihao Cen.

In this work performed in the framework of the PhD thesis of Zhihao Cen, and published as an Inria report RR-7868 [29], We consider a model of medium-term commodity contracts management. Randomness takes place only in the prices on which the commodities are exchanged, whilst state variable is multi-dimensional, and decision variable is integer. In our previous article, we proposed an algorithm based on the quantization of random process and a dual dynamic programming type approach to solve the continuous relaxation problem. In this paper, we study the multi-stage stochastic mixed integer linear program (SMILP) and show the difficulty when using dual programming type algorithm. We propose an approach based on the cutting plane method combined with the algorithm in our previous article, which gives an upper and a lower bound of the optimal value and a sub-optimal integer solution. Finally, a numerical test on a real problem in energy market is performed.

6.3.2. Two methods of pruning Benders' cuts and their application to the management of a gas portfolio

Participant: Laurent Pfeiffer.

This report, coauthored with R. Apparigliato and S. Auchapt (Gdf Suez), and published as Inria report 8133 [31], describes a gas portfolio management problem, which is solved with the SDDP (Stochastic Dual Dynamic Programming) algorithm. We present some improvements of this algorithm and focus on methods of pruning Benders' cuts, that is to say, methods of picking out the most relevant cuts among those which have been computed. Our territory algorithm allows a quick selection and a great reduction of the number of cuts. Our second method only deletes cuts which do not contribute to the approximation of the value function, thanks to a test of usefulness. Numerical results are presented.

6.4. Hamilton-Jacobi approach

6.4.1. Hamilton-Jacobi equations in singular domains

Participants: Zhiping Rao, Hasnaa Zidani.

A good deal of attention has been devoted to the analysis of Hamilton–Jacobi equations adapted to unconventional domains, particularly in view of application to control problems and traffic models. The topic is new and capable of interesting developments, the results so far obtained have allowed to clarify under reasonable assumptions, basic items as the right notion of viscosity solution to be adopted and the validity of comparison principles.

- The work [19], co-authored with C. Imbert (LAMA, U. Paris-Est) and R. Monneau (Cermics, ENPC), focuses on a Hamilton-Jacobi approach to junction problems with applications to traffic flows. More specifically, the paper is concerned with the study of a model case of first order Hamilton-Jacobi equations posed on a *junction*, that is to say the union of a finite number of half-lines with a unique common point. The main result is a comparison principle. We also prove existence and stability of solutions. The two challenging difficulties are the singular geometry of the domain and the discontinuity of the Hamiltonian. As far as discontinuous Hamiltonians are concerned, these results seem to be new. They are applied to the study of some models arising in traffic flows. The techniques developed here provide new powerful tools for the analysis of such problems.

- This work deals with deterministic control problems where the dynamic can be completely different in multi-complementary domains of the space IR^d. As a consequence, the dynamics present discontinuities at the interfaces of these domains. This leads to a complex interplay that has to be analyzed among transmission conditions to "glue" the propagation of the value function on the interfaces. Several questions arise: how to define properly the value function and what is the right Bellman Equation associated to this problem?. In the case of finite horizon problems without runing cost, a jonction condition is derived on the interfaces, and a precise viscosity notion is provided in a paper in progress. Moreover, a uniqueness result of a viscosity solution is shown.

6.4.2. A general Hamilton-Jacobi framework for nonlinear state-constrained control problems

Participants: Olivier Bokanowski, Hasnaa Zidani.

This work [10], co-authored with Albert Altarovici, deals with deterministic optimal control problem with state constraints and nonlinear dynamics. It is known for such a problem that the value function is in general discontinuous and its characterization by means of an HJ equation requires some controllability assumptions involving the dynamics and the set of state constraints. Here, we first adopt the viability point of view and look at the value function as its epigraph. Then, we prove that this epigraph can always be described by an auxiliary optimal control problem free of state constraints, and for which the value function is Lipschitz continuous and can be characterized, without any additional assumptions, as the unique viscosity solution of a Hamilton-Jacobi equation. The idea introduced in this paper bypasses the regularity issues on the value function of the constrained control problem and leads to a constructive way to compute its epigraph by a large panel of numerical schemes. Our approach can be extended to more general control problems. We study in this paper the extension to the infinite horizon problem as well as for the two-player game setting. Finally, an illustrative numerical example is given to show the relevance of the approach.

6.4.3. State-constrained optimal control problems of impulsive differential equations

Participants: Nicolas Forcadel, Zhiping Rao, Hasnaa Zidani.

The research report [35] presents a study on optimal control problems governed by measure driven differential systems and in presence of state constraints. The first result shows that using the graph completion of the measure, the optimal solutions can be obtained by solving a reparametrized control problem of absolutely continuous trajectories but with time-dependent state-constraints. The second result shows that it is possible to characterize the epigraph of the reparametrized value function by a Hamilton-Jacobi equation without assuming any controllability assumption

6.4.4. Level-set approach for reachability analysis of hybrid systems under lag constraints

Participants: Giovanni Granato, Hasnaa Zidani.

The study in [36] aims at characterizing a reachable set of a hybrid dynamical system with a lag constraint in the switch control. The setting does not consider any controllability assumptions and uses a level-set approach. The approach consists in the introduction of an adequate hybrid optimal control problem with lag constraints on the switch control whose value function allows a characterization of the reachable set. The value function is in turn characterized by a system of quasi-variational inequalities (SQVI). We prove a comparison principle for the SQVI which shows uniqueness of its solution. A class of numerical finite differences schemes for solving the system of inequalities is proposed and the convergence of the numerical solution towards the value function is studied using the comparison principle. Some numerical examples illustrating the method are presented. Our study is motivated by an industrial application, namely, that of range extender electric vehicles. This class of electric vehicles uses an additional module *the range extender* as an extra source of energy in addition to its main source a high voltage battery. The methodolgy presented in [36] is used to establish the maximum range of a Hybrid vehicle, see [22].

6.5. Collision avoidance and motion planning

6.5.1. Collision analysis for a UAV

Participants: Anna Désilles, Hasnaa Zidani.

The *Sense and Avoid* capacity of Unmanned Aerial Vehicles (UAV) is one of the key elements to open the access to airspace for UAVs. In order to replace a pilot's *See and Avoid* capacity such a system has to be certified "as safe as a human pilot on-board". The problem is to prove that an unmanned aircraft equipped with a S&A system can comply with the actual air transportation regulations. A paper in progress aims to provide mathematical and numerical tools to link together the safety objectives and sensors specifications. Our approach starts with the natural idea of a specified "safety volume" around the aircraft: the safety objective is to guarantee that no other aircraft can penetrate this volume. We use a general reachability and viability concepts to define nested sets which are meaningful to allocate sensor performances and manoeuvring capabilities necessary to protect the safety volume. Using the general framework of HJB equations for the optimal control and differential games, we give a rigorous mathematical characterization of these sets. Our approach allows also to take into account some uncertainties in the measures of the parameters of the incoming traffic. We also provide numerical tools to compute the defined sets, so that the technical specifications of a S&A system can be derived in accordance with a small set of intuitive parameters. We consider several dynamical models corresponding to the different choices of maneuvers (lateral, longitudinal and mixed). Our numerical simulations show clearly that the nature of used maneuvers is an important factor in the specifications of sensor's performances.

6.6. Numerical methods for HJ equations

6.6.1. An adaptive sparse grid semi-lagrangian scheme for first order Hamilton-Jacobi Bellman equations

Participant: Olivier Bokanowski.

The paper [14], co-authored with M. Griebel (Fraunhofer SCAI & Univ. Bonn), J. Garcke and I. Klopmpaker (TUB, Berlin) proposes a semi-Lagrangian scheme using a spatially adaptive sparse grid to deal with non-linear time-dependent Hamilton-Jacobi Bellman equations. We focus in particular on front propagation models in higher dimensions which are related to control problems. We test the numerical efficiency of the method on several benchmark problems up to space dimension $d = 8$, and give evidence of convergence towards the exact viscosity solution. In addition, we study how the complexity and precision scale with the dimension of the problem.

6.6.2. A discontinuous Galerkin scheme for front propagation with obstacles

Participant: Olivier Bokanowski.

In [33], co-authored with C.-W. Shu (Brown Univ.) and Y. Cheng (Michigan Univ.), some front propagation problems in the presence of obstacles are analysed. We extend a previous work (Bokanowski, Cheng and Shu, SIAM J. Scient. Comput., 2011), to propose a simple and direct discontinuous Galerkin (DG) method adapted to such front propagation problems. We follow the formulation of (Bokanowski, Forcadel and Zidani, SIAM J. Control Optim. 2010), leading to a level set formulation driven by $\min(u_t + H(x, \nabla u), u - g(x)) = 0$, where $g(x)$ is an obstacle function. The DG scheme is motivated by the variational formulation when the Hamiltonian H is a linear function of ∇u, corresponding to linear convection problems in the presence of obstacles. The scheme is then generalized to nonlinear equations, written in an explicit form. Stability analysis is performed for the linear case with Euler forward, a Heun scheme and a Runge-Kutta third order time discretization using the technique proposed in (Zhang and Shu, SIAM J. Control and Optim., 2010). Several numerical examples are provided to demonstrate the robustness of the method. Finally, a narrow band approach is considered in order to reduce the computational cost.

6.6.3. Semi-Lagrangian discontinuous Galerkin schemes for some first and second order PDEs

Participant: Olivier Bokanowski.

Explicit, unconditionally stable, high order schemes for the approximation of some first and second order linear, time-dependent partial differential equations (PDEs) are proposed in [34], in collaboration with G. Simarmata (internship 2011, currently in RI dep. of Rabobank). The schemes are based on a weak formulation of a semi-Lagrangian scheme using discontinuous Galerkin elements. It follows the ideas of the recent works of Crouseilles, Mehrenberger and Vecil (2010) and of Qiu and Shu (2011), for first order equations, based on exact integration, quadrature rules, and splitting techniques. In particular we obtain high order schemes, unconditionally stable and convergent, in the case of linear second order PDEs with constant coefficients. In the case of non-constant coefficients, we construct "almost" unconditionally stable second order schemes and give precise convergence results. The schemes are tested on several academic examples, including the Black and Scholes PDE in finance.

CONCHA Project-Team

6. New Results

6.1. Convergence of adaptive finite element algorithms

Participants: Roland Becker, Shipeng Mao, David Trujillo.

The theoretical analysis of mesh-adaptive methods is a very active field of research. We have generalized our previous results concerning optimality of adaptive methods to nonconforming finite elements [53]. Our results include the error due to iterative solution of the system matrices by means of a simple stopping criterion related to the error estimator. The main difficulty was the treatment of the nonconformity which leads to a perturbation of the orthogonality relation at the heart of the proofs for conforming finite elements. We have been able to extend this result to the Stokes equations, considering different lowest-order nonconforming finite elements on triangular and quadrilateral meshes [16].

In [19] we have shown that the smallness assumption required in all former proofs of optimality of adaptive finite element methods can be overcome, at least in some situations.

Finally, we have shown optimality of a new goal-oriented method in [21].

Our theoretical studies, which are motivated by the aim to develop better adaptive algorithms, have been accompanied by software implementation with the Concha library, see Section 5.1 . It hopefully opens the door to further theoretical and experimental studies.

6.2. Finite element methods for interface problems

Participants: Nelly Barrau, Roland Becker, Robert Luce.

The original formulation of NXFEM [63] is based on the doubling of elements. In some situations, as the case of a moving interface, it is computationally more convenient to have a method with local enrichment, as for the standard XFEM. In [47] we have developed such an approach based on NXFEM. We have developed an hierarchical formulation for a fictitious domain formulation in [7].

One of the technical difficulties is the simultaneous robustness of the method with respect to the size of the intersection of a mesh cell with the interface and with respect to the discontinuous diffusion parameters. In [ ] (note CRAS 2012) we proposed a modified formulation of the NXFEM which allows us to obtain this robustness to solve the Darcy equation.

In connection with the thesis of Nelly Barrau, supervised by Robert Luce and Eric Dubach (LMAP) we have:

- implemented lots of geometrical tools in 2D and 3D necessary to use the NXFEM methods,
- extended the method to P_k and Q_k finite elements ([42],
- generalized the residual estimator and developed an adaptative process with hanging node (8),
- adapted the method to the transport equation.

6.3. A posteriori error estimators based on $H(\mathrm{div})$-reconstructed fluxes

Participants: Roland Becker, Daniela Capatina, Robert Luce.

Mesh adaptivity is nowadays an essential tool in numerical simulations; in order to achieve it, reliable and efficient, easily computable *a posteriori* error estimators are needed. Such estimators obtained by reconstructing locally conservative fluxes in the Raviart-Thomas finite element space have been largely employed in the past years.

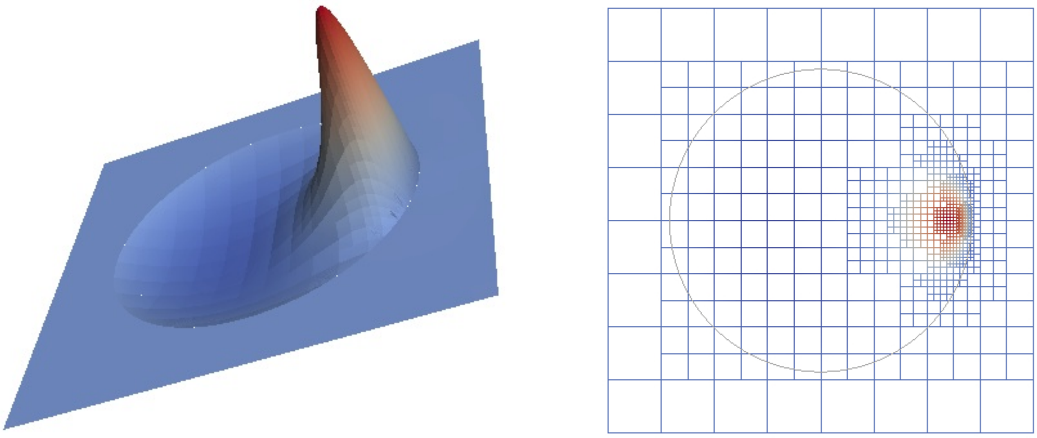


Figure 8. Result of an adaptative process with hanging node

We have so far considered the convection-diffusion equation and proposed a unified framework for several finite element approximations (conforming, nonconforming and discontinuous Galerkin). The main advantage of our approach is to use, contrarily to the existing references, only the primal mesh for the flux reconstruction, which presents certain facilities from a computational point of view.

For this purpose, the construction of the $H(\mathrm{div})$-vector involved in the error estimator is inspired by the hypercircle method cf. [56] and is achieved on patches, which may overlap. A patch depends on the type of the employed finite elements and is defined as the support of a basis function.

Our first results were presented in [12]. We are working on the extension to higher-order approximations, to quadrilateral meshes and to other model problems.

6.4. Discretization of Euler's equations

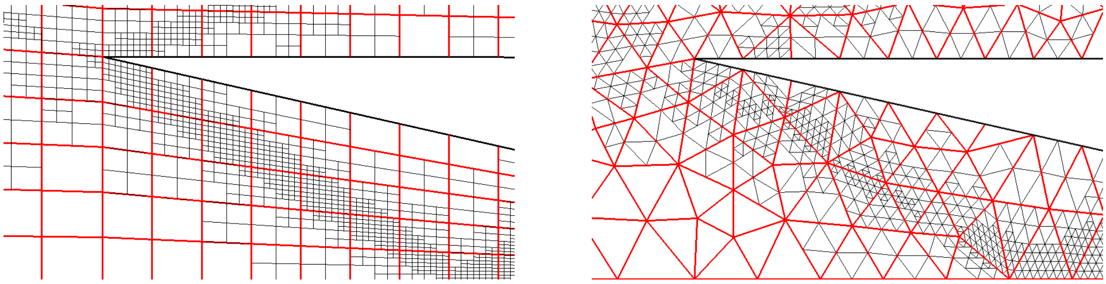
Participants: Roland Becker, Kossivi Gopki, Eric Schall, David Trujillo.

Over the past years, significant advances have been made in developing discontinuous Galerkin finite element methods (DGFEM) for applications in fluid flow and heat transfer. Certain features of the method have made it attractive as an alternative to other popular methods such as finite volume and more convenient finite element methods in thermal fluid engineering analyses. The DGFEM has been used successfully to solve hyperbolic systems of conservation laws. It makes use of the same local function space as the continuous method, but with relaxed continuity at inter-element boundaries. Since it uses discontinuous piecewise polynomial bases, the discretization is locally conservative and in the considered lowest-order case, the method preserves the maximum principle for scalar equations.

One of the challenges in Computational Fluid Dynamic (CFD) is to obtain as accurate as possible the solution of the problem under consideration at very low cost in terms of computational time. So our principal work is to find some relevant and robust strategies and technics of meshes adaptation in order to concentrate just the calculation where there are physical phenomena to capture. From Industrial point of view, the aim is to get the stationary solution as quick as possible with as much accuracy as possible. The main limitation of these results in CFD concern the underlying models: for example, nearly nothing seems to be known for (even linear) first-order systems or for realistic nonlinear equations. We therefore have developed different modern techniques, especially adaptive methods, to tackle this kind of problems in compressible CFD. The strategy is to iteratively improve the quality of the approximate solutions based on computed information (a posteriori error analysis).

In this way, a sequence of locally refined meshes is constructed, which allows for better efficiency as compared to more classical approaches in the presence of different kind of singularities. The main goal is to improve the aerodynamical design process for complex configurations by significantly reducing the time from geometry to solution at engineering-required accuracy using high-order adaptive methods.

One of our strategies of refinement is based on the creation of hanging nodes commonly called non-conforming refinement. The figures 9 show superposition of two kinds of meshes. One is a non-conforming refined mesh (black color) and the other one is the initial grid (red color) on which the refinement has been performed. It shows the technic of cutting the cells where singularities occur in the scramjet inlet.



(a) (b)

Figure 9. Superposition of non-conforming adapted black color) grid and initial grid (red color) – (a) quadrangles and (b) triangles.

The mesh adaptation is designed using some criteria as a posteriori error estimates. We have designed criteria based on the calculation of the jump of physical quantities like density, pressure, entropy, temperature and mach number at the inter-element. This criteria seems to be a very good indicator for the mesh adaptation. Figure 10 is the comparison of isoline of the density in scramjet internal flow at mach 3 of the initial mesh, the third and the sixth mesh after refinement.The indicator used is the density jump. It shows the impact and the accuracy of the solution obtained after the sixth iteration of the refinement.

The figure 11 shows the streamlines of the density in the scramjet inlet after the seventh iteration. This shows how the adaptation depicts almost clearly and accurately the shock waves and the expansion waves and their interactions in the domain.

Figure 12 represent the density isolines of a flow past cylinder test case using the non-conforming mesh adaptation with quadrangular an triangular girds.

We have also settled another indication which is hierarchical. It measures the difference of g_h with the physical quantity $g_{h/2}$ obtained by computation on a globally refined mesh $h/2$. This allows us the make comparison with the previous indicator. The case test considered for this comparison is an external flows past a cylinder airfoil at fixed free stream conditions : $M_\infty = 3$. The result is quite surprising the way one type of indicator can capture phenomenon that are not capture by the another one. In fact the hierarchical indicator seems to capture recirculation downstream to the obstacle which was not capture by the jump indicator (see figure 13)

We compare the computational time between a non-conforming mesh refinement and a globally mesh refined with nearly the same amount of cells. The meshes contain quadrangles or triangles. We can observe trough the following tables that the adapted meshes wether triangular or quadrangular meshes allow to save 20 to 90 times the computational time than the normal globally refined mesh. (see tables 1 and 2)

In table 1, the gain in time is 35 times in quadrangular grid case and 90 times triangular ones and in table 2, the gain in time: 18 times in quadrangular grid case and 58 times triangular ones. So one can say that the

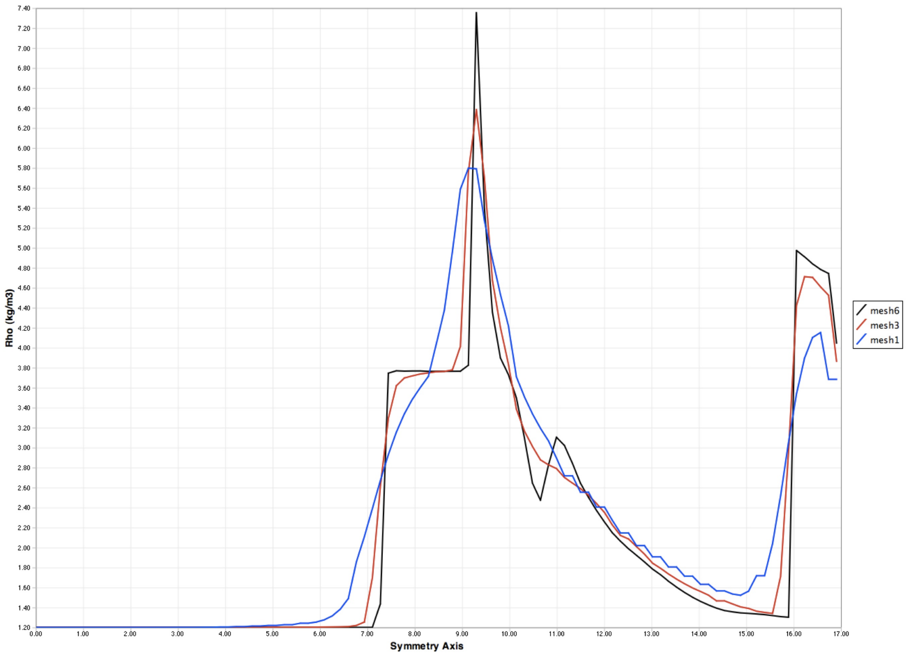


Figure 10. Cutlines along the symmetry axis of various meshes for the scramjet test case

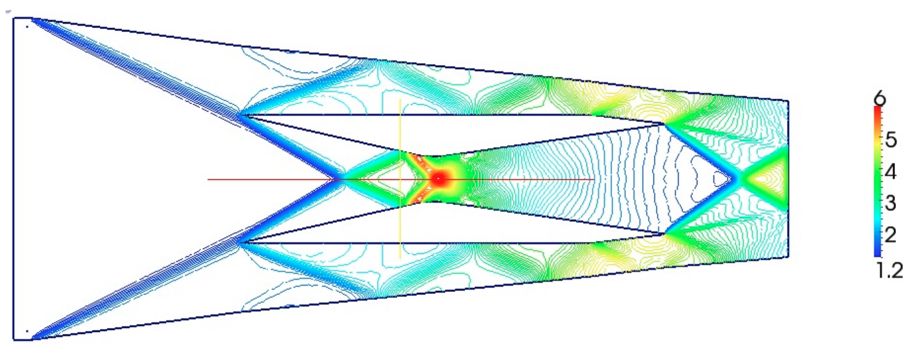
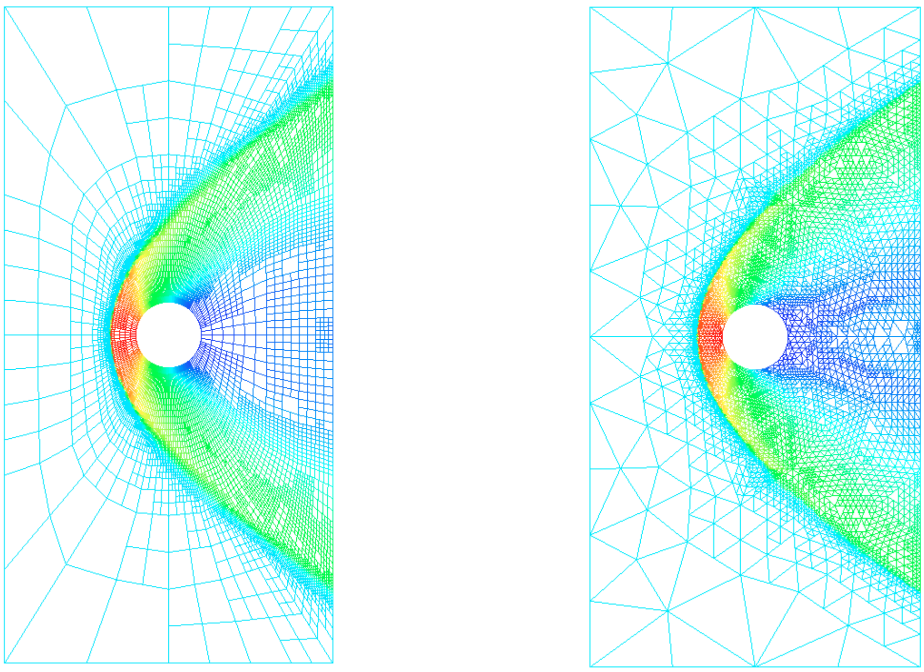
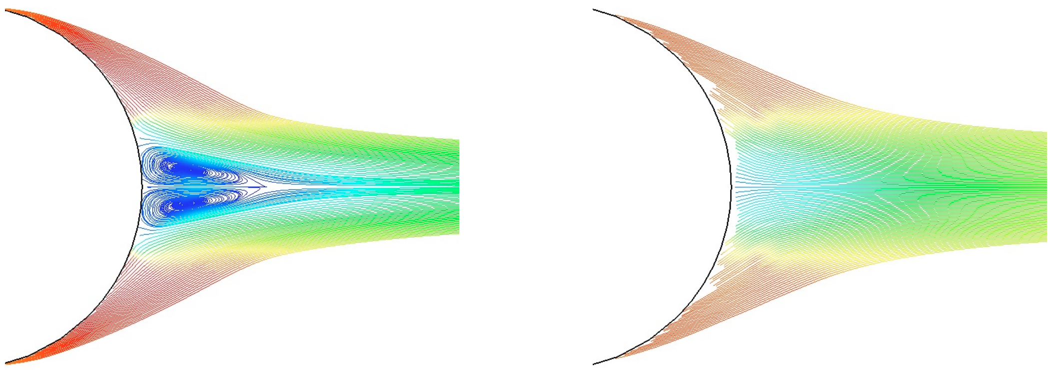


Figure 11. Density streamlines on grid obtained after the seveneth iteration of adaptive refinement procedure with density jump as indicator



(a) (b)

Figure 12. Locally adapted mesh on quadrilaterals (a) and triangles (b)



(a) (b)

Figure 13. Streamlines coloured by the density on meshes generated with hierarchical indicator (a) and with jump indicator (b)

Scramjet test case at mach=3

	Nodes	Cells	Segments	Compt. Time(s)
Scram_Quad_4	17043	15485	34308	25.0236
Scram_Quad_Uniform	17183	16640	33824	865.0177
Scram_Tri_4	9951	17005	29138	22.3141
Scram_Tri_Uniform	13295	25504	38800	2000.4269

Table 1

Flow past cylinder test case at mach=3

	Nodes	Cells	Segments	Compt. Time(s)
Cyl_Quad_5	11203	10174	23105	47.2187
Cyl_Quad_Uniform	10496	10240	20736	814.6168
Cyl_Tri_6	6480	10867	19264	79.7836
Cyl_Tri_Uniform	6032	11776	17808	4258.6618

Table 2

Figure 14. Comparison of computational times

adaptive mesh with the strategies and technics we have settled are efficient and robust in capturing physical phenomenon at a very reasonable low cost.

In concluding, the procedure of refinement permit to save computational time and have good accuracy of the approximated solution computed. Our focus is to continue the improve our methods and strategies in order to meet the requirement of accuracy, robustness and efficiency. Many other works are in hand such as slope limiters for high-order Discontinous Galerkin, low mach number computation with some remarkable approaches.

CORIDA Project-Team

6. New Results

6.1. Analysis and control of fluids and of fluid-structure interactions

In [38], a new characteristics method for the discretization of the two dimensional fluid-rigid body problem is proposed in the case of different densities for the fluid and the solid. Convergence results are obtained for a fully-discrete finite element scheme.

In [35], controllability results are obtained for a low Reynolds number swimmer. The swimmer is undergoing radial and axi-symmetric deformations in order to propel itself in a viscous fluid.

The aim of the paper [51] is to tackle the time optimal controllability of an (n+1)-dimensional nonholonomic integrator. A full description of the optimal control and optimal trajectories are explicitly obtained.

In [25], we the interaction between a viscous incompressible fluid and an elastic structure immersed in the fluid.

In [30], weconsider the model composed by a rigid body immersed into a n incompressible perfect fluid and analyze the regularity of the trajectory of the rigid body and of the fluid particles.

In [39], we study the motion of a rigid body with a cavity filled with a viscous liquid.

In [34], we analyze a model of vesicle moving into a viscous incompressible fluid.

In [27], we obtain the identifiability of a rigid body moving in a stationary viscous fluid.

In [40] we study a mathematical model for the dynamics of vesicle membranes in a 3D incompressible viscous fluid. we show that, given $T > 0$, for initial data which are small (in terms of T), these solutions are defined on $[0, T]$ (almost global existence).

6.2. Frequency domain methods for the analysis and control of systems governed by PDE's

In [21] and [20], we propose an asymptotic analysis for the simple layer potential for multiple scattering at low frequencies.

In [19] we propose some strategies to solve numerically the difficult problem of multiple scattering by a large number of disks at high frequency. To achieve this, we combine a Fourier series decomposition with the EFIE integral equation. Numerical examples will be presented to show the efficiency of our method.

In [32], we are concerned with the convergence analysis of the iterative algorithm for solving initial data inverse problems from partial observations that has been recently proposed in Ramdani et al. More precisely, we provide a complete numerical analysis for semi-discrete (in space) and fully discrete approximations derived using finite elements in space and an implicit Euler method in time. The analysis is carried out for abstract Schrödinger and wave conservative systems with bounded observation (locally distributed).

In [23], we propose a strategy to determine the Dirichlet-to-Neumann (DtN) operator for infinite, lossy and locally perturbed hexagonal periodic media, using a factorization of this operator involving two non local operators. The first one is a DtN type operator and corresponds to a half-space problem, while the second one is a Dirichlet-to-Dirichlet (DtD) type operator related to the symmetry properties of the problem.

In [18], we investigate absorbing boundary conditions for the two-dimensional Schrödinger equation with a time and space varying exterior potential.

6.3. Observality, controllability and stabilzation in the time domain

In [17] we consider N Euler-Bernoulli beams and N strings alternatively connected to one another and forming a chain beginning with a string. We study the strong and polynomial stabilities of this system on this network and the spectrum of the corresponding conservative system.

In [37] we study the asymptotic behavior of the solution of the non-homogeneous elastic systems with voids and a thermal effect. Our main results concern strong and polynomial stabilities (since this system suffers of exponential stability).

In [12], we consider the approximation of two coupled wave equations with internal damping. Our goal is to damp the spurious high frequency modes by introducing a numerical viscosity term in the approximation schemes and prove the exponential or polynomial decay of the discrete scheme.

In [13], we show similar results as in [12] for an abstract second order evolution equations.

In [44] we consider a class of infinite dimensional systems involving a control function u taking values in $[0, 1]$ and we prove, when u is given in an appropriate feedback form and the system satisfies appropriate observability assumptions, that the system is weakly stable. The main example concerns the analysis and stabilization of a model of Boost converter connected to a load via a transmission line.

In [46] we present a course on stabilization of hyperbolic equations given at a CIME session on Control of PDE's in Italy in July 2010, including well-known results, together with recent ones including nonlinear stabilization, memory-damping and stabilization of coupled systems by a reduced number of controls. In particular, we present the optimal-weight convexity method (Alabau-Boussouira 2005, 2010) in both the finite dimensional and infinite dimensional framework and give applications to semi-discretization of hyperbolic PDE's.

In [41], we consider stabilization of coupled systems of wave-type, with localized couplings and either localized internal closed loop controls or boundary control. We establish polynomial decay rates for coupling and damping regions which do not intersect in the one-dimensional case. We also derive results in the multi-dimensional case, under multiplier type conditions for both the coupling and damping regions. The novelty and difficulty is to consider localized couplings.

In [15], we give a constructive proof of Gibson's stability theorem, some extension and further positive and negative applications of this result.

In [36] we prove that the boundary controls for the heat equation have the bang-bang property, at least in rectangular domains. This result is proved by combining methods from traditionally distinct fields: the Lebeau-Robbiano strategy for null controllability and estimates of the controllability cost in small time for parabolic systems, on one side, and a Remez-type inequality for Muntz spaces and a generalization of Turan's inequality, on the other side.

In [16] we prove exact controllability for symmetric coupled wave equations by a single control in the case of coupling and control regions which do not intersect. For this, we use and extend the two-level energy method introduced by Alabau-Boussouira (2001, 2003). Using transmutation, we derive null controllability results for coupled parabolic and Schrödinger equations. This is the first positive quantitative result, in a multi-dimensional framework with control and coupling regions with empty intersection.

In [14], we prove controllability results for abstract systems of weakly coupled N evolution equations in cascade by a reduced number of boundary or locally distributed controls ranging from a single up to N-1 controls. We give applications to cascade coupled systems of N multi-dimensional hyperbolic, parabolic and diffusive (Schrödinger) equations. The results are valid for control and coupling regions which do not necessarily intersect.

In [22], we study two notions of controllability, called respectively radial controllability and directional controllability. We prove that for families of linear vector fields, the two notions are actually equivalent.

In [24] we solve an optimization problem in convex geometry which, despite its seeming simplicity, offers a nice variety of solutions, some of them being unexpectable.

The paper [28] is devoted to prove that the union of two identical balls minimizes a non linear eigenvalue (related to the generalized Wirtinger inequality) among sets of given volume.

In [33] is considered a problem in population dynamics where we investigate the question of optimal location of the zone of control.

In [26], we give a rigorous proof, valid also for unbounded operators, of the widely used "rotating wave approximations" for bilinear Schrödinger equations.

In [42], we exploit the results of [26] on standard examples of bilinear quantum systems.

CQFD Project-Team

5. New Results

5.1. Singularly Perturbed Discounted Markov Control Processes in a General State Space

Participant: François Dufour.

Markov decision processes, optimal control, infinite discounted expected cost, optimal control, singular perturbation

In this work, it is studied the asymptotic optimality of discrete-time Markov Decision Processes (MDP's in short) with general state space and action space and having weak and strong interactions. The idea in this work is to consider a MDP with general state and action spaces and to reduce the dimension of the state space by considering an averaged model. This formulation is often described by introducing a small parameter $\epsilon > 0$ in the definition of the transition kernel, leading to a singularly perturbed Markov model with two time scales. Our objective is twofold. First it is shown that the value function of the control problem for the perturbed system converges to the value function of a limit averaged control problem as ϵ goes to zero. In the second part of this work, it is proved that a feedback control policy for the original control problem defined by using an optimal feedback policy for the limit problem is asymptotically optimal. Our work extends existing results of the literature in the following two directions: the underlying MDP is defined on general state and action spaces and we do not impose strong conditions on the recurrence structure of the MDP such as Doeblin's condition.

These results have been obtained in collaboration with Oswaldo Luis Do Valle Costa from Escola Politécnica da Universidade de São Paulo, Brazil.

It has been published in SIAM Journal of Control and Optimization [16].

5.2. The expected total cost criterion for Markov decision processes under constraints: a convex analytic approach.

Participant: François Dufour.

Markov decision process, expected total cost criterion, constraints, linear programming, occupation measure

This work deals with discrete-time Markov Decision Processes (MDP's) under constraints where all the objectives have the same form of an expected total cost over the infinite time horizon. The existence of an optimal control policy is discussed by using the convex analytic approach. We work under the assumptions that the state and action spaces are general Borel spaces and the model is non-negative, semi-continuous and there exists an admissible solution with finite cost for the associated linear program. It is worth noting that, in contrast with the classical results of the literature, our hypotheses do not require the MDP to be transient or absorbing. Our first result ensures the existence of an optimal solution to the linear program given by an occupation measure of the process generated by a randomized stationary policy. Moreover, it is shown that this randomized stationary policy provides an optimal solution to this Markov control problem. As a consequence, these results imply that the set of randomized stationary policies is a sufficient set for this optimal control problem. Finally, our last main result states that all optimal solutions of the linear program coincide on a special set with an optimal occupation measure generated by a randomized stationary policy. Several examples are presented to illustrate some theoretical issues and the possible applications of the results developed in the paper.

These results have been obtained in collaboration with Alexey Piunovskiy from Department. of Mathematical Sciences, The University of Liverpool, United Kingdom and with Masayuki Horiguchi form the Department of Mathematics, Faculty of Engineering, Kanagawa University, Japan.

It has been published in Advances in Applied Probability [17] and in the invited session of the 25th conference EURO 2012 [27]

5.3. Approximation of Infinite Horizon Discounted Cost Markov Decision Processes

Participant: François Dufour.

Markov decision processes, infinite horizon discounted cost criterion, approximation and discretization

In this work, we deal with a discrete-time infinite horizon Markov decision process with locally compact Borel state and action spaces, and possibly unbounded cost function. Based on Lipschitz continuity of the elements of the control model, we propose a state and action discretization procedure for approximating the optimal value function and an optimal policy of the original control model. We provide explicit bounds on the approximation errors.

These results have been obtained in collaboration with Tomas Prieto-Rumeau, Department of Statistics and Operations Research, UNED, Madrid, Spain.

It has been published in the book Optimization, Control, and Applications of Stochastic Systems. In Honor of Onésimo Hernandez-Lerma [52].

5.4. Continuous Control of Piecewise Deterministic Markov Processes with Long Run Average Cost

Participant: François Dufour.

Piecewise-deterministic Markov Processes, long-run average cost, optimal control, integro-differential optimality equation

The main goal of this work is to derive sufficient conditions for the existence of an optimal control strategy for the long run average continuous control problem of piecewise deterministic Markov processes (PDMP's) taking values in a general Borel space and with compact action space depending on the state variable. In order to do that we apply the so-called vanishing discount approach to obtain a solution to an average cost optimality inequality (ACOI) associated to the long run average cost problem. Our main assumptions are written in terms of some integro-differential inequalities related to the so-called expected growth condition, and geometric convergence of the post-jump location kernel associated to the PDMP.

These results have been obtained in collaboration with Oswaldo Luis Do Valle Costa from Escola Politécnica da Universidade de São Paulo, Brazil.

It has been published in the book Stochastic Processes, Finance and Control. A Festschrift in Honor of Robert J. Elliott [51].

5.5. Optimal stopping for partially observed piecewise-deterministic Markov processes

Participants: Adrien Brandejsky, Benoîte de Saporta, François Dufour.

We have investigated an optimal stopping problem under partial observation for piecewise-deterministic Markov processes (PDMP) both from the theoretical and numerical points of view. PDMP's have been introduced by Davis [73] as a general class of stochastic models. They form a family of Markov processes involving deterministic motion punctuated by random jumps. One important property of a PDMP, relevant for the approach developed in this paper, is that its distribution is completely characterized by the embedded discrete time Markov chain $(Z_n, S_n)_{n\in\mathbb{N}}$ where Z_n is the n-th post-jump location and S_n is the n-th inter-jump time. We consider the following optimal stopping problem for a partially observed PDMP $(X_t)_{t\geq 0}$. Roughly speaking, the observation process $(Y_t)_{t\geq 0}$ is a point process defined through the embedded discrete time Markov chain $(Z_n, S_n)_{n\in\mathbb{N}}$. The inter-arrival times are given by $(S_n)_{n\in\mathbb{N}}$ and the marks by a noisy function of $(Z_n)_{n\in\mathbb{N}}$. For a given reward function g and a computation horizon $N \in \mathbb{N}$, we study the following optimal stopping problem

$$\sup_{\sigma \leq T_N} \mathbb{E}\left[g(X_\sigma)\right],$$

where T_N is the N-th jump time of the PDMP $(X_t)_{t\geq 0}$, σ is a stopping time with respect to the natural filtration $\mathcal{F}^o = (\mathcal{F}^o_t)_{t\geq 0}$ generated by the observations $(Y_t)_{t\geq 0}$.

A general methodology to solve such a problem is to split it into two sub-problems. The first one consists in deriving the filter process given by the conditional expectation of X_t with respect to the observed information $\mathcal{F}^o_t$. Its main objective is to transform the initial problem into a completely observed optimal stopping problem where the new state variable is the filter process. The second step consists in solving this reformulated problem, the new difficulty being its infinite dimension. Indeed, the filter process takes values in a set of probability measures.

Our work is inspired by [92] which deals with an optimal stopping problem under partial observation for a Markov chain with finite state space. The authors study the optimal filtering and convert their original problem into a standard optimal stopping problem for a continuous state space Markov chain. Then they propose a discretization method based on a quantization technique to approximate the value function. However, their method cannot be directly applied to our problem for the following main reasons related to the specificities of PDMPs.

Firstly, PDMPs are continuous time processes. Then, it appears natural to work with the embedded Markov chain $(Z_n, S_n)_{n\in\mathbb{N}}$. In addition, we assume that $(Z_n)_{n\in\mathbb{N}}$ takes finitely many values. However, an important difficulty is that the structure of stopping time remains intrinsically continuous. Consequently, our problem cannot be converted into a fully discrete time problem.

Secondly, the distribution of a PDMP combines both absolutely continuous and singular components. This is due to the existence of forced jumps when the process hits the boundary of the state space. As a consequence the derivation of the filter process is not straightforward. In particular, the absolute continuity hypothesis **(H)** of [92] does not hold.

Thirdly, in our context the reformulated optimization problem is not standard, unlike in [92]. Indeed, although we obtain a reformulation similar to an optimal stopping problem for a fully observed PDMP, it involves the Markov chain $(\Pi_n, S_n)_{n\in\mathbb{N}}$ that is not the embedded Markov chain of some PDMP. Therefore, a new derivation of dynamic programming equations is required as we cannot use the results of [81]. In particular, one needs to derive fine properties of the structure of the $(\mathcal{F}^o_t)_{t\geq 0}$-stopping times. Moreover, we construct an ϵ-optimal stopping time.

Finally, a natural way to proceed with the numerical approximation is then to follow the ideas developed in [92] [8] namely to replace the filter Π_n and the inter-jump time S_n by some finite state space approximations in the dynamic programming equation. However, a noticeable difference from [8] lies in the fact that the dynamic programming operators therein were Lipschitz continuous whereas our new operators are only Lipschitz continuous between some points of discontinuity. We overcome this drawback by splitting the operators into their restrictions onto their continuity sets. This way, we obtain not only an approximation of the value function of the optimal stopping problem but also an ϵ-optimal stopping time with respect to the filtration $(\mathcal{F}^o_t)_{t\geq 0}$ that can be computed in practice.

This work is submitted for publication [60] and presented in an invited international conference [26].

5.6. Predictive maintenance for the heated hold-up tank

Participants: Benoîte de Saporta, François Dufour, Huilong Zhang.

A complex system is inherently sensitive to failures of its components. One must therefore determine maintenance policies in order to maintain an acceptable operating condition. Optimizing the maintenance is a very important problem in the analysis of complex systems. It determines when it is best that maintenance tasks should be performed on the system in order to optimize a cost function: either maximize a performance function or conversely minimize a loss function. Moreover, this optimization must take into account the random nature of failures and random evolution and dynamics of the system.

The example considered here is the maintenance of the heated hold-up tank, a well know test case for dynamic reliability, see e.g. [75], [89], [90], [94]. The system consists of a tank containing a fluid whose level is controlled by three components: two inlet pumps and one outlet valve. A thermal power source heats up the fluid. The failure rate of the components depends on the temperature, the position of the three components monitors the liquid level in the tank, and in turn, the liquid level determines the temperature. The main characteristic of this system is that it can be modeled by a stochastic hybrid process, where the discrete and continuous parts interact in a closed loop. As a consequence, simulating this process and computing related reliability indices has been a challenge for the dynamic reliability community. To our best knowledge, optimization of maintenance policies for the heated hold-up tank has not been addressed yet in the literature.

The only maintenance operation considered here is the complete replacement of all the failed components and the system restarts in its initial equilibrium state. Partial repairs are not allowed. Mathematically, this problem of preventive maintenance corresponds to a stochastic optimal stopping problem as explained by example in the book of Aven and Jensen [68]. It is a difficult problem because of the closed loop interactions between the state of the components and the liquid level and temperature. A classical approach consists in using condition-based maintenance (CBM) to act on the system based on its current state and before its failure. One can for example calculate the remaining useful life (RUL) of the system and the preventive replacement is carried out when the deterioration level exceeds a certain threshold or enters in a certain state [96], [80]. Our approach also takes into account the current state of the process, but our decision rule is not based on damage accumulation nor does it correspond to hitting some threshold. Instead, it involves a performance function that reflects that the longer the system is in a functioning state the better.

The dynamics of the heated hold-up tank can be modeled by a piecewise deterministic Markov process (PDMP), see [94]. Therefore, our maintenance problem boils down to an optimal stopping problem for PDMP's. PDMP's are a class of stochastic hybrid processes that has been introduced by Davis [73] in the 80's. These processes have two components: a Euclidean component that represents the physical system (e.g. temperature, pressure, ...) and a discrete component that describes its regime of operation and/or its environment. Starting from a state x and mode m at the initial time, the process follows a deterministic trajectory given by the laws of physics until a jump time that can be either random (e.g. it corresponds to a component failure or a change of environment) or deterministic (when a magnitude reaches a certain physical threshold, for example the pressure reaches a critical value that triggers a valve). The process restarts from a new state and a new mode of operation, and so on. This defines a Markov process. Such processes can naturally take into account the dynamic and uncertain aspects of the evolution of the system. A subclass of these processes has been introduced by Devooght [75] for an application in the nuclear field. The general model has been introduced in dynamic reliability by Dutuit and Dufour [79].

As illustrated above, it is crucial to have an efficient numerical tool to compute the optimal maintenance time in practical cases. To this aim, a general numerical approach was developed in [8]. It was first applied to an example of maintenance of a metallic structure subject to corrosion, without closed loop interactions or deterministic jumps, and with a simple cost function that did not depend on time, see [23]. The objective of the present paper is to further demonstrate the high practical power of the theoretical methodology described in [8], by applying it to the more challenging heated hold-up tank problem. The cost function chosen here is also more complex as it takes into account both continuous components as well as the running time. More precisely, we propose to compute the optimal cost as well as a quasi-optimal stopping rule, which is the date when the maintenance should be performed. As a by-product of our procedure, the distribution of the optimal maintenance dates is also obtained, as well as the distributions of the liquid level and temperature at the chosen maintenance date.

This work is submitted for publication [66] and presented in an international conference [32].

5.7. Efficient simulation of the availability of a feedwater control system

Participants: Benoîte de Saporta, François Dufour, Huilong Zhang.

In the reliability modeling of complex control systems, classical methodologies such as even-trees/fault-trees or Petri nets may not represent adequately the dynamic interactions existing between the physical processes (modeled by continuous variables) and the functional and dysfunctional behavior of its components (modeled by discrete variables). We have proposed a framework for modeling and simulation of a water level control system in the steam generator (SG) in the secondary circuit of a nuclear power plant. A similar benchmark system was described by the U.S. Nuclear Regulatory Commission [67] to compare two approaches to dynamic reliability: DFM (Dynamic Flowgraph Methodology) and Markov/CCMT (Cell-to-Cell Mapping Technique). But the report released by the NRC is not sufficient to reconstruct a realistic model. We have developed a complete benchmark case. The behavioral model of SG is obtained from a linearized model published in 2000 by EDF [87]. Detailed description of the components, failure modes and control laws of the principal components is presented. For modeling the system, we use the piecewise deterministic Markov processes (PDP) framework [73] and for implementation we chose Simulink associated with Stateflow. PDP's offer a very general modeling framework to deal with dynamic reliability problems; Simulink is a good tool to simulate non linear differential equations and their controller, while Stateflow implementation is appropriate for finite state machine descriptions of different components.

In our benchmark system, four physical processes are considered: feedwater flowrate, steam flow, narrow range water level and wide range water level. A PID controller is used to maintain the water level within limits of set-points. The system is composed of seven components: 1 passive system representing vapor transport system, 3 extraction pumps, 2 feeding turbopumps, and 1 waterflow regulation valve. The functional and dysfunctional behaviors and the failure rates of each component are based on operational experience. In 2012, we have further improved our simulator by taking captors (and their possible failures) into account.

This work was presented in an international conference [36], a national conference [39] and is published as a book chapter [49].

5.8. Stochastic control for underwater optimal trajectories

Participants: Benoîte de Saporta, François Dufour, Huilong Zhang.

This work aims to compute optimal trajectories for underwater vehicles evolving in a given environment to accomplish some tasks. This is an optimal control problem. In real context, available inputs are not perfectly known. Hence a stochastic approach seems to be needed. Markov decision processes (MDPs) constitute a general family of controlled stochastic processes suitable for the modeling of sequential decision-making problems. The analysis of MDPs leads to mathematical and computational problems. The corresponding theory has reached a rather high degree of maturity, although the classical tools (such as value iteration, policy iteration, linear programming, and their various extensions) are generally hardly applicable in practice. Hence, solving MDPs numerically is an awkward and important problem. The method is applied to control a submarine which wants to well detect one or several targets. Why? A smart operator, if provided information about target's position and velocity and a sound propagation code can find a good trajectory. If we- now consider a submarine surrounded by several targets, it is clear that a human operator will have great difficulty to find the best route.

This work was presented in an international conference [35].

5.9. Statistical study od asymmetry in cell lineage data

Participants: Benoîte de Saporta, Anne Gégout-Petit.

This work proposes a rigorous methodology to study cell division data consisting in several observed genealogical trees of possibly different shapes. For instance, [93] filmed 94 colonies of Escherichia coli cells dividing between four and nine times. We propose a new rigorous approach to take into account all the available information. Indeed, we propose an inference based on a finite fixed number of replicated trees when the total number of observed cells tends to infinity. We use the missing data asymmetric BAR model introduced by [7]. In this approach, the observed genealogies are modeled with a two-type Galton Watson (GW) process. However, we propose a different least-squares estimator for the parameters of the BAR process that does not correspond to the single-tree estimators averaged on the replicated trees. We also propose an estimator of the parameters of the GW process specific to our binary tree structure and not based simply on the observation of the number of cells of each type in each generation.

Our procedure allows us to fully take into account missing observations, data from different trees as well as the dependence structure within genealogical trees. It also enables us to use all the information available without the drawbacks of low accuracy for estimators or low power for tests on small single trees. We study the consistency and asymptotic normality of our estimators and derive asymptotic confidence intervals as well as Wald's type tests to investigate the asymmetry of the data for both the BAR and GW processes. Our results are applied to the Escherichia coli data of [93].

This work is in collaboration with Laurence Marsalle (Lille 1 University). It is submitted for publication [65] and was presented in an international conference [33].

5.10. Random coefficient bifurcating autoregressive processes

Participants: Benoîte de Saporta, Anne Gégout-Petit.

In the 80's, Cowan and Staudte [72] introduced Bifurcating Autoregressive processes (BAR) as a parametric model to study cell lineage data. A quantitative characteristic of the cells (e.g. growth rate, age at division) is recorded over several generations descended from an initial cell, keeping track of the genealogy to study inherited effects. As a cell usually gives birth to two offspring by division, such genealogies are naturally structured as binary trees. BAR processes are thus a generalization of autoregressive processes (AR) to this binary tree structure, by modeling each line of descent as a first order AR process, allowing the environmental effects on sister cells to be correlated. Statistical inference for the parameters of BAR processes has been widely studied, either based on the observation of a single tree growing to infinity [72], [85], [83], [95] or on a large number of small independent trees [86], [84].

Various extensions of the original model have been proposed, but to our best knowledge, only two papers [71] and [70] deal with random coefficient BAR processes. In the former by Bui and Huggins it is explained that random coefficients BAR processes can account for observations that do not fit the usual BAR model. For instance, the extra randomness can model irregularities in nutrient concentrations in the media in which the cells are grown. In this work, we propose a new model for random coefficient BAR processes (R-BAR). It is more general than that of Bui and Huggins, as the random variables are not supposed to be Gaussian, they may not have moments of all order and correlation between all the sources of randomness are allowed. Moreover, we propose an asymmetric model in the continuance of [82], [69], [74], [70], [7], [24] in the context of missing data. Indeed, experimental data are often incomplete and it is important to take this phenomenon into account for the inference. We model the structure of available data by a Galton Watson tree, instead of a complete binary tree. Our model is close to that developed in [70], but the assumptions on the noise process are different as we allow correlation between the two sources of randomness but require higher moments because of the missing data and because we do not use a weighted estimator. The main difference is that the model in [70] is fully observed, whereas ours allows for missing observations.

Our approach for the inference of our model is also different from [71], [70]. As we cannot use maximum likelihood estimation, we propose modified least squares estimators as in [91]. The originality of our approach is that it combines the bifurcating Markov chain and martingale approaches. Bifurcating Markov chains (BMC) were introduced in [82] on complete binary trees and further developed in [74] in the context of missing data on Galton Watson trees. BAR models can be seen as a special case of BMC. This interpretation allows us to establish the convergence of our estimators. A by-product of our procedure is a new general result for BMC on Galton Watson trees. Indeed, in [82], [74] the driven noise sequence is assumed to have moments of all order. Here, we establish new laws of large numbers for polynomial functions of the BMC where the noise sequence only has moments up to a given order. The strong law of large numbers [78] and the central limit theorem for martingales have been previously used in the context of BAR processes and adapted to special cases of martingales on binary trees. In this paper, we establish a general law of large numbers for square integrable martingales on Galton Watson binary trees. This result is applied to our R-BAR model to obtain sharp convergence rates and a quadratic strong law for our estimators.

This work is in collaboration with Laurence Marsalle (Lille 1 University). It is submitted for publication [64].

5.11. Hidden Markov Model for the detection of a degraded state in an optronic equipment

Participants: Camille Baysse, Anne Gégout-Petit, Jérôme Saracco.

As part of optimizing the reliability, Thales Optronics now includes systems that examine the state of its equipment. This function is performed by HUMS (Health & Usage Monitoring System). We hope to implement a program based on these observations that can determine the lifetime of this optronic equipment. Our study focuses on a simple example of HUMS. As part of our research, we are interested in a variable called "time-to cold" noted TMF, which reflects the state of system. Using this information about this variable, we seek to detect as soon as possible a degraded state and propose maintenance before failure. For this we use a hidden Markov model. The state of our system at time t is then modeled by a Markov chain X_t. However we do not observe directly this chain but indirectly through the TMF, a noisy function of this chain. Thanks to filtering equations, we obtained results on the probability that an equipment breaking down at time t, knowing the history of the TMF until this moment. We have subsequently studied this methodology with simulated data. Then finally we applied these results on the analysis of our real data and we have checked that the results are consistent with the reality. So using this method could allow the company to recall equipments which are estimated in deteriorated state and do not control those estimated in stable state. Thales Optronics could improve its maintenance system and reduce its cost function.

This work is a part of the CIFRE PhD of Camille Baysse also supervised for the Thales part by Didier Bihannic and Michel Prenat. It was presented in an national conference [38] and is submitted for publication in an national per-reviewed journal [58].

5.12. Predictive maintenance for an optronic equipment

Participants: Camille Baysse, Benoîte de Saporta, Anne Gégout-Petit, Jérôme Sarraco.

After the problem of detection of a degraded state, we have tackled the problem of predictive maintenance for an optronic equipment. For this we model the state of the system by a PDMP (state with three possible values and cumulative time of use). In this framework, we reformulate the problem of maintenance of optimization in an optimal stopping problem maximizing a criteria about time of use without failure. In this framework, we can use known results developed in the CQFD team on optimal control [8], [23]. We have extensively studied the problem with simulated data, computed grid of quantization and optimal policy for the real problem. This results will be implemented by Thales in HUMS of optronic equipment.

This work was presented in an national conference [38] and an abstract is accepted for publication in an international conference with papers.

5.13. Non parametric estimation of the jump rate for non-homogeneous marked renewal processes

Participants: Romain Azaïs, François Dufour, Anne Gégout-Petit.

This work is devoted to the nonparametric estimation of the jump rate and the cumulative rate for a general class of non-homogeneous marked renewal processes, defined on a separable metric space. In our framework, the estimation needs only one observation of the process within a long time. Our approach is based on a generalization of the multiplicative intensity model, introduced by Aalen in the seventies. We provide consistent estimators of these two functions, under some assumptions related to the ergodicity of an embedded chain and the characteristics of the process. The methodology is illustrated by a numerical example. It is the object of a paper [57] to appear in the Annales de l'Intitut Poincaré

5.14. Non parametric estimation of conditional distribution of the interjumping times for piecewise Markov processes

Participants: Romain Azaïs, François Dufour, Anne Gégout-Petit.

This work gives a nonparametric method for estimating the conditional density associated to the jump rate of a piecewise-deterministic Markov process. In our framework, the estimation needs only one observation of the process within a long time interval. Our method relies on a generalization of Aalen's multiplicative intensity model. We prove the uniform consistency of our estimator, under some reasonable assumptions related to the primitive characteristics of the process. A simulation example illustrates the behavior of our estimator. This work is the object of a paper [56] submitted for publication

5.15. Stochastic modelling and simulation of fatigue crack propagation using piecewise-deterministic Markov processes

Participants: Romain Azaïs, Anne Gégout-Petit.

Fatigue crack propagation is a stochastic phenomenon in nature due to the inherent uncertainties coming from material properties, environmental conditions and loads. Stochastic processes offer an appropriate framework for modelling crack propagation since it is intended to include sources variabilities. In this work, we propose to model crack propagation mechanism with Piecewise Deterministic Markov Process (PDMP) using usual random crack laws. Conventional laws proposed in the literature seem inadequate for describing the whole fatigue crack trajectory mainly when the crack extends in a rapid manner. To overcome this drawback, a new modelling is proposed that consists in using more than one law as each one is more suitable for a specific phase during crack propagation. Regime-switching models seem very attractive and with our modelling assessed crack growth rates and crack lengths are very close to experimental values. Moreover, behaviour just before failure is well captured and can be discussed. Empirical curves from literature are used to adjust the parameters associated to the proposed modelling. Statistical observations and numerical simulations show the efficiency of the proposed approach to model and to simulate fatigue crack growth. This work has been presented in an international congress [34] and is the object of a paper which will be submitted very soon.

5.16. Statistical Analysis of Grapevine Mortality Associated with Esca or Eutypa Dieback Foliar Expression

Participant: Anne Gégout-Petit.

Esca and Eutypa dieback are two major wood diseases of grapevine in France. Their widespread distribution in vineyards leads to vine decline and to a loss in productivity. However, little is known either about the temporal dynamics of these diseases at plant level, and equally, the relationships between foliar expression of the diseases and vine death is relatively unknown too. . To investigate these questions, we surveyed the vines of six vineyards cv. Cabernet Sauvignon in the Bordeaux region, by recording foliar symptoms, dead arms and dead plants from 2004 to 2010. In 2008, 2009 and 2010, approximately five percent of the asymptomatic vines died but the percentage of dead vines which had previously expressed esca foliar symptoms was higher, and varied between vineyards. A logistic regression model was used to select the previous years of symptomatic expression associated with vine mortality. The mortality of esca is always associated with the foliar symptom expression of the year preceding vine death. One or two other earlier years of expression frequently represented additional risk factors. The Eutypa dieback symptom was also a risk factor of death, superior or equal to that of esca. The study of the internal necroses of vines expressing esca or Eutypa dieback is discussed in the light of these statistical results. This work has been presented in an international congress [44] and is the object of a submitted paper.

5.17. MonteCarlo test for two patterns of point processes on a grid

Participants: Anne Gégout-Petit, Marie Chavent, Amaury Labenne.

In order to compare two patterns of distribution of symptomatic or dead vines in a same vineyard but for two consecutive years, we have developed a Monte Carlo test. First we estimate the intensity of occurrence of disease in one of the pattern, then we simulate n realizations i.i.d. of this intensity and compute the associate likelihoods in order to build an interval that cover $(1-\alpha)$ per cent of the realizations. The test reject the equality of repartition if the likelihood computed with the second pattern is not included in this interval. We have made simulations and applied this test to the repartition of esca in vineyard. This work has been presented in a national workshop on software R [46].

5.18. Multivariate Analysis for the detection of the effect of a treatment

Participant: Anne Gégout-Petit.

The aim of this work is to give some statistical rules to determine if a patient is meeting a given treatment (a BD here). The criterium commonly used to determine if a patient is meeting a BD treatment is based only on one physiological parameter : if this parameter increases, the patient is meeting. But now, many physiological parameters are measured in routine and it seems that a patient could have a global amelioration of his health state due to the treatment without an increase of the single used parameter.

Using standard multivariate analysis techniques, and classification, we have proposed criteria to discriminate groups of patients different in regard of their response to treatment. This work will be used by physiologists to propose new criteria for the measure of the effect of a BD treatment. It is in collaboration with physiologists from Bordeaux and Nantes universities and is the object of a submitted paper in a international peer-reviewed journal in the domain of pneumology .

5.19. A hidden renewal model for monitoring aquatic systems biosensors

Participants: Romain Azaïs, Raphaël Coudret.

This work aims at modeling signals of oysters' openings over time using a four-state renewal process. Two of them are of particular interest and correspond to instants when the animals are open or closed. An estimator of the cumulative jump rate of this process is provided. It relies on observations of the jumps between the four states. Here these measures are not available but the observed signal takes ranges of real values according to this underlying process. A procedure to estimate a probability density function that summarizes the information of the signal is explained. This leads to estimate the hidden renewal process and then its cumulative jump rate for each oyster. A classification of these functions for a group of oysters discriminate them according to their assumed health status. Such a diagnosis is essential when using these animals as biosensors for water quality assessment. This work is a joint work with Gilles Durrieu from Université de Bretagne Sud and in collaboration with UMR CNRS 5805 EPOC.

5.20. A recursive nonparametric estimator for the transition kernel of a piecewise-deterministic Markov process

Participant: Romain Azaïs.

We investigate a nonparametric approach to provide a recursive estimator of the transition density of a non-stationary piecewise-deterministic Markov process, from only one observation of the path within a long time. In this framework, we do not observe a Markov chain with transition kernel of interest. Fortunately, one may write the transition density of interest as the ratio of the invariant distributions of two embedded chains of the process. Our method consists in estimating these invariant measures. We state a result of consistency under some general assumptions about the main features of the process. A simulation study illustrates the well asymptotic behavior of our estimator. This work is the object of a paper [55] submitted for publication.

5.21. A new sliced inverse regression method for multivariate response

Participants: Jérôme Saracco, Raphaël Coudret.

We consider a semiparametric regression model of a q-dimensional multivariate response y on a p-dimensional covariate x. In this paper, a new approach is proposed based on sliced inverse regression for estimating the e ffective dimension reduction (EDR) space without requiring a prespeci ed parametric model. The convergence at rate square root of n of the estimated EDR space is shown. We discuss the choice of the dimension of the EDR space. The numerical performance of the proposed multivariate SIR method is illustrated on a simulation study. Moreover, we provide a way to cluster components of y related to the same EDR space. One can thus apply properly multivariate SIR on each cluster instead of blindly applying multivariate SIR on all components of y. An application to hyperspectral data is provided.

These results have been obtained in collaboration with Stéphane Girard (Inria Rhône Alpes).

The paper is under revision for possible publication in CSDA [63].

5.22. Comparison of kernel density estimators with assumption on number of modes

Participants: Jérôme Saracco, Raphaël Coudret.

A data-driven bandwidth choice for a kernel density estimator called critical bandwidth is investigated. This procedure allows the estimation to have as many modes as assumed for the density to estimate. Both Gaussian and uniform kernels are considered. For the Gaussian kernel, asymptotic results are given. For the uniform kernel, an argument against these properties is mentioned. These theoretical results are illustrated with a simulation study which compare the kernel estimators that rely on critical bandwidth with another one which uses a plug-in method to select its bandwidth. An estimator that consists in estimates of density contour clusters and takes assumptions on number of modes into account is also considered. Finally, the methodology is illustrated using environment monitoring data.

These results have been obtained in collaboration with Gilles Durrieu (Université Bretagne-Sud).

The paper is under revision for possible publication in Communications in Statistics - Simulation and Computation [62].

5.23. A new approach on recursive and non-recursive SIR methods

Participant: Jérôme Saracco.

We consider a semiparametric single index regression model involving a p-dimensional quantitative covariable x and a real dependent variable y. A dimension reduction is included in this model via an index $x'\beta$. Sliced inverse regression (SIR) is a well-known method to estimate the direction of the Euclidean parameter β which is based on a ”slicing step” of y in the population and sample versions. The goal of this paper is twofold. On the one hand, we focus on a recursive version of SIR which is also suitable for multiple indices model. On the other hand, we propose a new method called SIRoneslice when the regression model is a single index model. The SIRoneslice estimator of the direction of β is based on the use of only one ”optimal” slice chosen among the H slices. Then, we provide its recursive version.We give an asymptotic result for the SIRoneslice approach. Simulation study shows good numerical performances of the SIRoneslice method and clearly exhibits the main advantage of using recursive versions of the SIR and SIRoneslice methods from a computational time point of view. A real dataset is also used to illustrate the approach. Some extensions are discussed in concluding remarks. The proposed methods and criterion have been implemented in R and the corresponding codes are available from the authors.

These results have been obtained in collaboration with Bernad Bercu (Université Bordeaux 1) and Thi Mong Ngoc Nguyen (Université de Strasbourg).

The paper has been published in the Journal of the Korean Statistical Society [11].

5.24. On the asymptotic behavior of the Nadaraya-Watson estimator associated with the recursive SIR method

Participant: Jérôme Saracco.

We investigate the asymptotic behavior of the Nadaraya-Watson estimator for the estimation of the regression function in a semiparametric regression model. On the one hand, we make use of the recursive version of the sliced inverse regression method for the estimation of the unknown parameter of the model. On the other hand, we implement a recursive Nadaraya-Watson procedure for the estimation of the regression function which takes into account the previous estimation of the parameter of the semiparametric regression model.We establish the almost sure convergence as well as the asymptotic normality for our Nadaraya-Watson estimator. We also illustrate our semiparametric estimation procedure on simulated data.

These results have been obtained in collaboration with Bernad Bercu (Université Bordeaux 1) and Thi Mong Ngoc Nguyen (Université de Strasbourg) .

The paper is submitted [59].

5.25. Comparison of sliced inverse regression approaches for underdetermined cases

Participants: Jérôme Saracco, Raphaël Coudret.

Among methods to analyze high-dimensional data, the sliced inverse regression (SIR) is of particular interest for non-linear relations between the dependent variable and some indices of the covariate. When the dimension of the covariate is greater than the number of observations, classical versions of SIR cannot be applied. Various upgrades were then proposed to tackle this issue such as RSIR and SR-SIR, to estimate the parameters of the underlying model and to select variables of interest. In this paper, we introduce two new estimation methods respectively based on the QZ algorithm and on the Moore-Penrose pseudo-inverse. We also describe a new selection procedure of the most relevant components of the covariate that relies on a proximity criterion between submodels and the initial one. These approaches are compared with RSIR and SR-SIR in a simulation study. Finally we applied SIR-QZ and the associated selection procedure to a genetic dataset in order to find eQTL.

These results have been obtained in collaboration with Benoit Liquet (Université Bordeaux 2) . The paper is submitted .

5.26. Orthogonal rotation in PCAMIX

Participants: Marie Chavent, Jérôme Saracco.

Kiers (1991) considered the orthogonal rotation in PCAMIX, a principal component method for a mixture of qualitative and quantitative variables. PCAMIX includes the ordinary Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA) as special cases. In this work, we give a new presentation of PCAMIX where the principal components and the squared loadings are obtained from a Singular Value Decomposition. The loadings of the quantitative variables and the principal coordinates of the categories of the qualitative variables are also obtained directly. In this context, we propose a computationaly efficient procedure for varimax rotation in PCAMIX and a direct solution for the optimal angle of rotation. A simulation study shows the good computational behavior of the proposed algorithm. An application on a real data set illustrates the interest of using rotation in MCA. All source codes are available in the R package “PCAmixdata”.

These results have been obtained in collaboration with Vanessa Kuentz of IRSTEA (UR ABDX).

It has been published in Advances in Data Analysis and Classification [15] and presented in the context of application in cultural sociology in the Premières Rencontres R [42].

5.27. A sliced inverse regression approach for data stream

Participants: Marie Chavent, Jérôme Saracco.

In this work, we focus on data arriving sequentially by block in a stream. A semiparametric regression model involving a common EDR (Effective Dimension Reduction) direction is assumed in each block. Our goal is to estimate this direction at each arrival of a new block. A simple direct approach consists in pooling all the observed blocks and estimate the EDR direction by the SIR (Sliced Inverse Regression) method. But some disadvantages appear in practice such as the storage of the blocks and the running time for high dimensional data. To overcome these drawbacks, we propose an adaptive SIR estimator of based on the SIR approach for a stratified population developed by Chavent et al. (2011). The proposed approach is faster both from computational complexity and running time points of view, and provides data storage benefits. We show the consistency of our estimator at the root-n rate and give its asymptotic distribution. We propose an extension to multiple indices model. We also provide a graphical tool in order to detect if a drift occurs in the EDR direction or if some aberrant blocks appear in the data stream. In a simulation study, we illustrate the good numerical behavior of our estimator. One important advantage of this approach is its adaptability to changes in the underlying model. Finally we apply it on real data concerning the estimation of Mars surface physical properties.

This work is under revision in Statistics and Computing [61].

5.28. ClustOfVar: An R Package for the Clustering of Variables

Participants: Marie Chavent, Jérôme Saracco.

Clustering of variables is as a way to arrange variables into homogeneous clusters, i.e., groups of variables which are strongly related to each other and thus bring the same information. These approaches can then be useful for dimension reduction and variable selection. Several specific methods have been developed for the clustering of numerical variables. However concerning qualitative variables or mixtures of quantitative and qualitative variables, far fewer methods have been proposed. The R package ClustOfVar was specifically developed for this purpose. The homogeneity criterion of a cluster is defined as the sum of correlation ratios (for qualitative variables) and squared correlations (for quantitative variables) to a synthetic quantitative variable, summarizing “as good as possible” the variables in the cluster. This synthetic variable is the first principal component obtained with the PCAMIX method. Two clustering algorithms are proposed to optimize the homogeneity criterion: iterative relocation algorithm and ascendant hierarchical clustering. We also propose a bootstrap approach in order to determine suitable numbers of clusters. We illustrate the methodologies and the associated package on small datasets.

These results have been obtained in collaboration with Vanessa Kuentz of IRSTEA (UR ABDX).

It has been published in Journal of Statistical Softwares [14]. The study of the inclusion of environment by the farmer with ClustOfVar has been presented in the Premières Rencontres R and in [45]

5.29. Divisive Monothetic Clustering for Interval and Histogram-valued Data

Participant: Marie Chavent.

In this paper we propose a divisive top-down clustering method designed for interval and histogram-valued data. The method provides a hierarchy on a set of objects together with a monothetic characterization of each formed cluster. At each step, a cluster is split so as to minimize intra-cluster dispersion, which is measured using a distance suitable for the considered variable types. The criterion is minimized across the bipartitions induced by a set of binary questions. Since interval-valued variables may be considered a special case of histogram-valued variables, the method applies to data described by either kind of variables, or by variables of both types. An example illustrates the proposed approach.

These results have been obtained in collaboration with Paula Brito of Porto University and presented in ICPRAM’2012 [31].

5.30. Classification of EEG signals by an evolutionary algorithm

Participants: Marie Chavent, Laurent Vézard.

The goal is to predict the alertness of an individual by analyzing the brain activity through electroencephalographic data (EEG) captured with 58 electrodes. Alertness is characterized as a binary variable that can be in a normal or relaxed state. We collected data from 44 subjects before and after a relaxation practice, giving a total of 88 records. After a pre-processing step and data validation, we analyzed each record and discriminate the alertness states using our proposed slope criterion. Afterwards, several common methods for supervised classification (k nearest neighbors, decision trees -CART-, random forests, PLS and discriminant sparse PLS) were applied as predictors for the state of alertness of each subject. The proposed slope criterion was further refined using a genetic algorithm to select the most important EEG electrodes in terms of classification accuracy. Results shown that the proposed strategy derives accurate predictive models of alertness.

These results have been obtained in collaboration with Pierrick Legrand of ALEA Inria team.

It has been published in Journal des Nouvelles Technologies [25] and presented in COMPSTAT 2012 [47].

5.31. Variable selection by genetic algorithm for the study of alertness states.

Participants: Marie Chavent, Laurent Vézard.

The aim of this work is to predict the state of alertness of an individual (binary variable, "normal" or "relaxed") from the study of brain activity (electroencephalographic signals EEG) collected with a limited number of electrodes. In fact, the set up of electrodes during the EEG signal acquisition is time consuming and these electrodes are correlated. In our study, the EEG of 58 participants in the two alertness states (116 records) were collected via a cap with 58 electrodes. After a data validation step based on the study of the contingent negative variation (CNV), 19 subjects were retained in the study. A CSP (Common Spacial Pattern) coupled to a linear discriminant analysis were used to build a decision rule and thus predict the alertness of the participants. A genetic algorithm was used to determine a subset of electrodes of size p '(where p' <p, where p = 58 is the number of electrodes). This presentation will present the CSP in the general framework and will introduce innovations made to this method. The genetic algorithm will be described proposed and recent results will be presented.

These results have been obtained in collaboration with Pierrick Legrand of ALEA Inria team.

It has been presented in the Journée Évolutionnaire Thématique, 23éme édition [48].

5.32. Handling Missing Values with Regularized Iterative Multiple Correspondence Analysis

Participant: Marie Chavent.

A common approach to deal with missing values in multivariate exploratory data analysis consists in minimizing the loss function over all non-missing elements. This can be achieved by EM-type algorithms where an iterative imputation of the missing values is performed during the estimation of the axes and components. This paper proposes such an algorithm, named iterative multiple correspondence analysis, to handle missing values in multiple correspondence analysis (MCA). This algorithm, based on an iterative PCA algorithm, is described and its properties are studied. We point out the over tting problem and propose a regularized version of the algorithm to overcome this major issue. Finally, performances of the regularized iterative MCA algorithm (implemented in the R-package named missMDA) are assessed from both simulations and a real dataset. Results are promising with respect to other methods such as the missing-data passive modi ed margin method, an adaptation of the missing passive method used in Gini's Homogeneity analysis framework.

It has been published in Journal of Classification [21].

DEFI Project-Team

6. New Results

6.1. Qualitative methods for inverse scattering problems

6.1.1. Sampling methods with time dependent data

Participant: Houssem Haddar.

Together with A. Lechleiter and S. Marmorat we proposed and analyzed a time domain linear sampling method as an algorithm to solve the inverse scattering problem of reconstructing an obstacle with Robin or Neumann boundary condition from time-dependent near-field measurements of scattered waves. Our algorithm is based on our earlier work to solve a similar inverse scattering problem for obstacles with Dirichlet boundary conditions. In addition to the analysis of a different scattering problem, we provided a substantial improvement of the method on both theoretical and numerical levels. More specifically, we analyzed the method for incident waves generated by pulses with bounded spectrum. Moreover, adapting the function space setting to this type of data allowed us to provide a simpler analysis. On the numerical side, we presented a fast implementation of the inversion algorithm that relies on a FFT-based evaluation of the near-field operator [34].

6.1.2. Inverse problems for periodic penetrable media

Participant: Dinh Liem Nguyen.

Imaging periodic penetrable scattering objects is of interest for non-destructive testing of photonic devices. The problem is motivated by the decreasing size of periodic structures in photonic devices, together with an increasing demand in fast non-destructive testing. In this project, we considered the problem of imaging a periodic penetrable structure from measurements of scattered electromagnetic waves. As a continuation of earlier work jointly with A. Lechleiter [24], [25], [23], we considered an electromagnetic problem for transverse magnetic waves (previous work treats transverse electric fields), and also the full Maxwell equations. In both cases, we treat the direct problem by a volumetric integral equation approach and construct a Factorization method [4], [44], [43], [48].

6.1.3. Transmission Eigenvalues and their application to the identification problem

Participant: Houssem Haddar.

The so-called interior transmission problem plays an important role in the study of inverse scattering problems from (anisotropic) inhomogeneities. Solutions to this problem associated with singular sources can be used for instance to establish uniqueness for the imaging of anisotropic inclusions from muti-static data at a fixed frequency. It is also well known that the injectivity of the far field operator used in sampling methods is related to the uniqueness of solutions to this problem. The frequencies for which this uniqueness fails are called transmission eigenvalues. We are currently developing approaches where these frequencies can be used in identifying (qualitative informations on) the medium properties. Our research on this topic is mainly done in the framework of the associate team ISIP http://www-direction.inria.fr/international/PHP/Networks/LiEA.php with the University of Delaware. A review article on the state of art concerning the transmission eigenvalue problem has been written in collaboration with F. Cakoni [32]. We are also in the process of editing a spacial issue of the journal Inverse Problems dedicated to the use of these transmission eigenvalues in inverse problems. Our recent contributions are the following:

- In collaboration with M. Fares and F. Collino from CERFACS and A. Cossonnière from ENSIEETA we finalized our work on the use of a surface integral equation approach to numerically compute transmission eigenvalues for inclusions with piecewise constant index. The main difficulty behind this procedure is the compactness of the obtained integral operator in usual Sobolev spaces associated with the forward scattering problem. We solved this difficulty by introducing a preconditioning operator associated with a “coercive” transmission problem. On the theoretical side, together with A; Cossonnière we also finalyzed the analysis of the Fredholm properties of the interior transmission problem for the cases where the index contrast changes sign outside the boundary by using the surface integral equation approach [16].

- With G. Giorgi, we developed a method that give estimates on the material properties using the first transmission eigenvalue. This method is based on reformulating the interior transmission eigenvalue problem into an eigenvalue problem for the material coefficients. We validated our methodology for homogeneous and inhomogeneous inclusions and backgrounds. We also treated the case of a background with absorption and the case of scatterers with multiple connected components of different refractive indexes [21].
- With F. Cakoni and D. Colton we initiated the study of transmission eigenvalues for absorbing media. In particular, we showed that, in the case of absorbing media, transmission eigenvalues form a discrete set, exist for sufficiently small absorption and for spherically stratified media exist without this assumption. For constant index of refraction, we also obtained regions in the complex plane where the transmission eigenvalues cannot exist and obtain a priori estimates for real transmission eigenvalues [14].
- With F. Cakoni and A. Cossonnière we considered the interior transmission problem corresponding to the inverse scattering by an inhomogeneous (possibly anisotropic) media in which an impenetrable obstacle with Dirichlet boundary conditions is embedded. Our main focus is to understand the associated eigenvalue problem, more specifically to prove that the transmission eigenvalues form a discrete set and show that they exist. The presence of Dirichlet obstacle brings new difficulties to already complicated situation dealing with a non-selfadjoint eigenvalue problem. In this work we employed a variety of variational techniques under various assumptions on the index of refraction as well as the size of the Dirichlet obstacle [15].

6.1.4. The factorization method for inverse scattering problems

6.1.4.1. The factorization method for cracks with impedance boundary conditions

Participants: Yosra Boukari, Houssem Haddar.

We use the Factorization method to retrieve the shape of cracks with impedance boundary conditions from farfields associated with incident plane waves at a fixed fre- quency. This work is an extension of the study initiated by Kirsch and Ritter [Inverse Problems, 16, pp. 89-105, 2000] where the case of sound soft cracks is considered. We address here the scalar problem and provide theoretical validation of the method when the impedance boundary conditions hold on both sides of the crack. We then deduce an inversion algorithm and present some validating numerical results in the case of simply and multiply connected cracks [38].

6.1.4.2. The factorization method for EIT with uncertain background

Participants: Giovanni Migliorati, Houssem Haddar.

We extended the Factorization Method for Electrical Impedance Tomography to the case of background featuring uncertainty. This work is based on our earlier algorithm for known but inhomogeneous backgrounds. We developed three methodologies to apply the Factorization Method to the more difficult case of piece-wise constant but uncertain background. The first one is based on a recovery of the background through an optimization scheme and is well adapted to relatively low dimensional random variables describing the background. The second one is based on a weighted combination of the indicator functions provided by the Factorization Method for different realiza- tions of the random variables describing the uncertain background. We show through numerical experiments that this procedure is well suited to the case where many real-izations of the measurement operators are available. The third strategy is a variant of the previous one when measurements for the inclusion-free background are available. In that case, a single pair of measurements is sufficient to achieve comparable accuracy to the deterministic case [42].

6.1.4.3. The factorization method for GIBC

Participants: Mathieu Chamaillard, Nicolas Chaulet, Houssem Haddar.

We are concerned with the identification of some obstacle and some Generalized Impedance Boundary Conditions (GIBC) on the boundary of such obstacle from far field measurements generated by the scattering of harmonic incident waves. The GIBCs are approximate models for thin coatings, corrugated surfaces, rough surfaces or imperfectly conducting media.

We justified the use of the Factorization method to solve the inverse obstacle problem in the presence of GIBCs. This method gives a uniqueness proof as well as a fast algorithm to reconstruct the obstacle from the knowledge of the far field produced by incident plane waves for all the directions of incidence at a given frequency. We also provided some numerical reconstructions of obstacles for several impedance operators.

6.2. Iterative Methods for Non-linear Inverse Problems

6.2.1. Inverse medium problem for axisymmetric eddy current models

Participants: Houssem Haddar, Zixian Jiang, Kamel Riahi.

We continued our developments of shape optimization methods for inclusion detection in an axisymmetric eddy current model. This problem is motivated by non-destructive testing methodologies for steam generators. We are finalizing our joint work with A. Lechleiter on numerical methods for the solution of the direct problem in weighted Sobolev spaces using approriate Dirichlet-to-Neumann mappings to bound the computational domain. We are also finalizing jointly with M. El Guedri the work on inverse solver using a regularized steepest descent method for the problem of identifying a magnetite deposits using axial eddy current probe.

We are currently investigating two research directions:

- The development of asymptotic models to identify thin highly conducting deposits. We derived three possible asymptotic models that can be exploited in the inverse problem. The numerical validation is under study.
- The extension of this work to 3D configurations with axisymmetric configuration at infinity, which has been started with the PostDoc of K. Riahi.

6.2.2. A min-max formulation for inverse scattering problems

Participants: Grégoire Allaire, Houssem Haddar, Dimitri Nicolas.

After having developed an inverse solver combining the use of Level-Set method and topological garadient method for multistatic inverse scattering problem and numerically showed how convergence can be achieved with intial guess provided by the Linear Sampling Method, we explored the use of an objective function that would lead to quicker and more stable reconstructions. This has been achieved through maximizing the least-square difference with respect to the Herglotz kernel of used incident wave while minimizing with respect to the geometrical parameters. Premliminary numerical experimentations showed that this procedure is viable and lead to quicker inversion algorithms [5].

6.2.3. The conformal mapping method and inverse scattering at low frequencies

Participant: Houssem Haddar.

Together with R. Kress we have employed a conformal mapping technique for the inverse problem to reconstruct a perfectly conducting inclusion in a homogeneous background medium from Cauchy data for electrostatic imaging, that is, for solving an inverse boundary value problem for the Laplace equation. In a recent work [41] we proposed an extension of this approach to inverse obstacle scattering for time-harmonic waves, that is, to the solution of an inverse boundary value problem for the Helmholtz equation. The main idea is to use the conformal mapping algorithm in an iterative procedure to obtain Cauchy data for a Laplace problem from the given Cauchy data for the Helmholtz problem. We presented the foundations of the method together with a convergence result and exhibit the feasibility of the method via numerical examples.

6.2.4. A steepest descent method for inverse electromagnetic scattering problems

Participants: Houssem Haddar, Nicolas Chaulet.

In a continuation of our earlier work jointly with L. Bourgeois [13], we studied the application of non linear optimization techniques to solve the inverse scattering problems for the $3D$ Maxwell's equations with generalized impedance boundary conditions. We characterized the shape derivative in the case where the GIBC is defined by a second order surface operator. We then applied a boundary variation method based on a regularized steepest descent to solve the 3-D inverse problem with partial farfield data. The obtained numerical results demonstrated the possibility of identifying the shape of coated objects as well as the parameters of the coating in the $3D$ Maxwell case.

6.3. Shape and topology optimization

6.3.1. Geometric constraints in shape and topology optimization

Participant: Grégoire Allaire.

With François Jouve (LJLL) and Georgios Michailidis (Renault and CMAP), we propose a method to handle geometric constraints in shape and topology optimization. In the framework of the level-set method we rely on a notion of local thickness which is computed using the signed-distance function to the boundary of the shape. We implement this method in two and three space dimensions for a model of linear elasticity. We consider various formulations of the constrained optimization problem and compute a shape derivative to advect the shape. We discuss different ways to handle the constraints. The resulting optimized shape is strongly dependent on the initial guess and on the way the constraints are being treated.

6.3.2. A hybrid optimization method

Participant: Grégoire Allaire.

With Charles Dapogny (Renault and LJLL) and Pascal Frey (LJLL) we propose a method for structural optimization that relies on two alternative descriptions of shapes : on the one hand, they are exactly meshed so that mechanical evaluations by finite elements are accurate ; on the other hand, we resort to a level-set characterization to describe their deformation along the shape gradient. The key ingredient is a meshing algorithm for building a mesh, suitable for numerical computations, out of a piecewise linear level-set function on an unstructured mesh. Therefore, our approach is at the same time a geometric optimization method (since shapes are exactly meshed) and a topology optimization method (since the topology of successive shapes can change thanks to the power of the level-set method).

6.3.3. DeHomogenization

Participant: Olivier Pantz.

In most shape optimization problems, the optimal solution does not belong to the set of genuine shapes but is a composite structure. The homogenization method consists in relaxing the original problem thereby extending the set of admissible structures to composite shapes. From the numerical viewpoint, an important asset of the homogenization method with respect to traditional geometrical optimization is that the computed optimal shape is quite independent from the initial guess (even if only a partial relaxation is performed). Nevertheless, the optimal shape being a composite, a post-treatment is needed in order to produce an almost optimal non-composite (i.e. workable) shape. The classical approach consists in penalizing the intermediate densities of material, but the obtained result deeply depends on the underlying mesh used and the details level is not controllable. We proposed in [51] a new post-treatment method for the compliance minimization problem of an elastic structure. The main idea is to approximate the optimal composite shape with a locally periodic composite and to build a sequence of genuine shapes converging toward this composite structure. This method allows us to balance the level of details of the final shape and its optimality. Nevertheless, it was restricted to particular optimal shapes, depending on the topological structure of the lattice describing the arrangement of the holes of the composite. We lifted this restriction in order to extend our method to any optimal composite structure for the compliance minimization problem in [50]. Since, the method has been improved and a new article presenting the last results is in preparation. Moreover, we intend to extend this approach to other kinds of cost functions. A first attempt, based on a gradient method, has been made. Unfortunately, it was leading to local minima. Thus a new strategy has to be worked out. It will be mainly based on the same ideas than the one developed for the compliance minimization problem, but some difficulties are still to be overcome.

6.3.4. Level-Set Method

Participant: Olivier Pantz.

We have begin to work, with Gabriel Delagado, on a new level-set optimization method, based on a gradient method. The key idea consists in computing directly the derivative of the discretized cost functions. The main advantage is that it is usually more simple to implement than the standard approach (consisting in using a discretized version of the gradient of the cost function). Moreover, the results obtained are as good or even better than the one obtained in previous works. Nevertheless, this method has its drawbacks, since the cost function is only derivable almost everywhere (the zero level-set has to be transverse to the triangulation of the mesh). It follows that convergence toward the minimum by the gradient method is not granted. To overcome this problem, we intend to use a mix-formulation for the state function. Unfortunately, such a formulation, in the case of linear elasticity is quite difficult to obtain. We thus intend to begin with the simplest scalar case, for which a lot more hybrid formulations are available.

6.3.5. Robust Optimization

Participant: Olivier Pantz.

One of the main problem in shape optimization problems is due to the fact that the gradient is never computed exactly. When the current solution is far from a local optimum, this is not a problem: even a rough approximation of the gradient enable us to exhibit a descent direction. On the contrary, when close to a local optimal, a very precise computation of the gradient is needed. We intend, with G. Delgado, to use a-posteriori error estimates evaluate the errors made on the computation of the gradient and to ensure that at each step, a genuine descent direction is used in the gradient method.

6.3.6. Level-set method applied to structural optimization with contact

Participants: Houssem Haddar, Olivier Pantz.

The current study covers the design and implementation of a method for topological shape optimization in order to optimize multi-connected structures taking into account the contact that may arise between the different components. This project is motivated by the optimization of leaf springs, issue proposed by the company CORTEL and is conducted in the framework of the Master internship of M. Mahjoub. We proposed a method that relies on the use of a Level Set Method coupled with a penalty method to handle contact with different components. The level set function is used for instance to construct the penalization functional. Preliminary results showed that the method efficiently handle optimal design with a targeted non linear deformation behavior prescribed by the manufacturer.

6.3.7. Optimization of a sodium fast reactor core

Participants: Grégoire Allaire, Olivier Pantz.

In collaboration with D. Schmidt, G. Allaire and E. Dombre, we apply the geometrical shape optimization method for the design of a SFR (Sodium Fast reactor) core in order to minimize a thermal counter-reaction known as the sodium void effect. In this kind of reactor, by increasing the temperature, the core may become liable to a strong increase of reactivity ρ, a key-parameter governing the chain-reaction at quasi-static states. We first use the 1 group energy diffusion model and give the generalization to the 2 groups energy equation. We then give some numerical results in the case of the 1 group energy equation. Note that the application of our method leads to some designs whose interfaces can be parametrized by very smooth curves which can stand very far from realistic designs. We don't explain here the method that it would be possible to use for recovering an operational design but there exists several penalization methods that could be employed to this end. This work was partially sponsored by EDF. Our results will be published in the proceedings of the CEMRACS'11, during which part of the results have been obtained.

6.4. Asymptotic Analysis

6.4.1. Asymptotic analysis of the interior transmission eigenvalues related to coated obstacles

Participants: Nicolas Chaulet, Houssem Haddar.

This work is a collaboration with Fioralba Cakoni from the University of Delaware (USA). The interior transmission eigenvalues play an important role in the area of inverse scattering problems. These eigenvalues can actually be determined by multi-static far field data. Thus, they could be used for non destructive testing. We focused on the case where the obstacle is a perfectly conducting body coated by some thin dielectric material. We derived and justified the asymptotic expansion of the first interior transmission eigenvalue with respect to the thickness of the coating for the TM electromagnetic polarization. This expansion provided interesting qualitative information about the behavior of these eigenvalues and also gave an explicit formula to compute the thickness of the coating.

6.4.2. Effective boundary conditions for thin periodic coatings

Participants: Mathieu Chamaillard, Houssem Haddar.

This topic is the object of a collaboration with Patrick Joly and is a continuation of our earlier work on interface conditions done in the framework of the PhD thesis of Berangère Delourme [18], [17]. Th goal here is to derive effective conditions that model scattering from thin periodic coatings where the thickness and the periodicity are of the same length but very small compared to the wavelength. The originality of our work, compared to abundant literature is to consider the case of arbitrary geometry (2-D or 3-D) and to consider higher order approximate models. We formally derived third order effective conditions after exhibiting the full asymptotic expansion of the solution in terms of the periodicity length.

6.4.3. Homogenization of thermal radiative transfer models in heterogeneous domains

Participant: Grégoire Allaire.

With my former PhD student, Zakaria Habibi, we studied the homogenization of heat transfer in periodic porous media where the fluid part is made of long thin parallel cylinders, the diameter of which is of the same order than the period. The heat is transported by conduction in the solid part of the domain and by conduction, convection and radiative transfer in the fluid part (the cylinders). A non-local boundary condition models the radiative heat transfer on the cylinder walls. To obtain the homogenized problem we first use a formal two-scale asymptotic expansion method. The resulting effective model is a convection-diffusion equation posed in a homogeneous domain with homogenized coefficients evaluated by solving so-called cell problems where radiative transfer is taken into account. In a second step we rigorously justify the homogenization process by using the notion of two-scale convergence. One feature of this work is that it combines homogenization with a 3D to 2D asymptotic analysis since the radiative transfer in the limit cell problem is purely two-dimensional. Eventually, we provide some 3D numerical results in order to show the convergence and the computational advantages of our homogenization method. We also focused on the contribution of the so-called second order corrector. If the source term is a periodically oscillating function (which is the case in our application to nuclear reactor physics), a strong gradient of the temperature takes place in each periodicity cell, corresponding to a large heat flux between the sources and the perforations. This effect cannot be taken into account by the homogenized model, neither by the first order corrector. We show that this local gradient effect can be reproduced if the second order corrector is added to the reconstructed solution. Z. Habibi received the 2012 Paul Caseau PhD prize in the field "modélisation et simulation numérique", prize created by the Académie des technologies and EDF.

6.4.4. Homogenization of complex flows in porous media

Participant: Grégoire Allaire.

With Robert Brizzi (CMAP), Jean-François Dufrêche (Marcoule and Montpellier), Andro Mikelic (Lyon 1) and Andrey Piatnitski (Narvik) we studied the homogenization (or upscaling) of a system of partial differential equations describing the non-ideal transport of a N-component electrolyte in a dilute Newtonian solvent through a rigid porous medium. Non-ideal effects are taken into account by the mean spherical approximation (MSA) model. We first study the existence of equilibrium solutions in the absence of external forces. When the motion is governed by a small static electric field and a small hydrodynamic force, we generalize O'Brien's argument to deduce a linearized model. We then proceed to the homogenization of these linearized equations and prove that the effective tensor satisfies Onsager properties, namely is symmetric positive definite. We eventually make numerical comparisons with the ideal case.

With my PhD student Harsha Hutridurga we study the convection and diffusion of a solute in a porous medium in the presence of a linear chemical reaction of adsorption/desorption on the pore surfaces. The mathematical model is a system of two coupled convection-diffusion equations, one in the bulk of the saturated fluid flowing in the porous medium, the other on the pore surface, at the interface with the solid part of the porous medium. The coupling takes place through a linear reaction term expressing the exchange of mass between the bulk concentration and the surface concentration. By a method of two-scale asymptotic expansion with drift we obtain the homogenized problem in a moving frame. We rigorously justify our upscaling approach by using the notion of two-scale convergence with drift. Some 2-d numerical tests are performed in order to study the effect of variations of the adsorption rate constant and surface molecular diffusion on the effective dispersion tensor.

With Irina Pankratova (Narvik) and Andrey Piatnitski (Narvik) we consider the homogenization of a non-stationary convection-diffusion equation posed in a bounded domain with periodically oscillating coefficients and homogeneous Dirichlet boundary conditions. Assuming that the convection term is large, we give the asymptotic profile of the solution and determine its rate of decay. In particular, it allows us to characterize the “hot spot”, i.e., the precise asymptotic location of the solution maximum which lies close to the domain boundary and is also the point of concentration. Due to the competition between convection and diffusion, the position of the “hot spot” is not always intuitive as exemplified in some numerical tests.

6.4.5. Multiscale finite elements

Participant: Grégoire Allaire.

With my PhD student Franck Ouaki we introduced a new multiscale finite element method to solve convection-diffusion problems where both velocity and diffusion coefficient exhibit strong variations at a much smaller scale than the domain of resolution. In that case, classical discretization methods, used at the scale of the heterogeneities, turn out to be too costly or useless. Our method aims at solving this kind of problems on coarser grids with respect to the size of the heterogeneities by means of particular basis functions. These basis functions are solutions to cell problems and are designed to reproduce the variations of the solution on an underlying fine grid. Since all cell problems are independent from each other, these problems can be solved in parallel, which makes the method very efficient when used on parallel architectures. The convergence proof of our method is still in progress. But, on the basis of results of periodic homogenization, an a priori error estimate, that represents a first step in the proof, has already been proved. A 2-d numerical implementation in FreeFem++ has also been performed.

6.4.6. A new shell modeling modeling

Participant: Olivier Pantz.

Using a formal asymptotic expansion, we have proved with K. Trabelsi, that non-isotropic thin-structure could behave (when the thickness is small) like a shell combining both membrane and bending effects. It is the first time to our knowledge that such a model is derived. An article on this is currently under review.

6.4.7. A new Liouville type Rigidity Theorem

Participant: Olivier Pantz.

We have recently developed a new Liouville type Rigidity Theorem. Considering a cylindrical shaped solid, we prove that if the local area of the cross sections is preserved together with the length of the fibers, then the deformation is a combination of a planar deformation and a rigid motion. The results currently obtained are limited to regular deformations and we are currently working with B. Merlet to extend them. Nevertheless, we mainly focus on the case where the conditions imposed to the local area of the cross sections and the length of the fibers are only "almost" fulfilled. This will enable us to derive rigorously new non linear shell models combining both membrane and flexural effects that we have obtained using a formal approach. An article on this subject is currently in preparation.

6.4.8. Lattices

Participant: Olivier Pantz.

With A. Raoult and N. Meunier (Université Paris Descartes), we have compute the asymptotic limit of a square lattice with three-points interactions. Considering such interaction is important in the case of square lattices, because such lattices, if only endowed with two-points closest neighbor interactions, show no resistance to compression, what is quit restrictive. We prove in particular that under some symmetry assumptions on the type of elementary interactions, no micro-relaxation do occur and that the limit can be obtained by a mere quasiconvexication. Without those assumptions, the computation of the limit requires the resolution of a homogenization problem on an infinite number of cells, what is usually out of reach. Our work has been published in M3AS [26].

6.5. Diffusion MRI

Participants: Jing-Rebecca Li, Houssem Haddar, Julien Coatléven, Dang Van Nguyen, Hang Tuan Nguyen.

Diffusion Magnetic Resonance Imaging (DMRI) is a promising tool to obtain useful information on microscopic structure and has been extensively applied to biological tissues. In particular, we would like to focus on two applications:

- inferring from DMRI measurements changes in the cellular volume fraction occurring upon various physiological or pathological conditions.

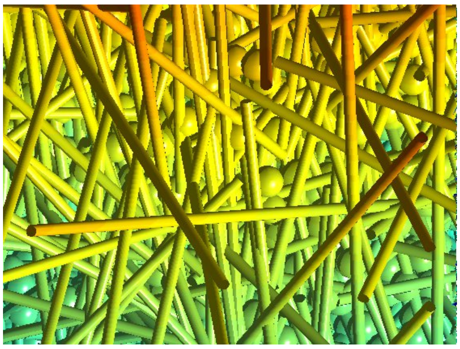


Figure 1. Computational domain for simulating diffusion in cerebral gray matter.

  This application is one of the first to show the promise of DMRI because it can detect acute cerebral ischemia (cell swelling) on the basis of lower than normal apparent diffusion coefficient a few minutes after stroke.

- estimating the average cell size in the case of tumor imaging

  This application is useful as a diagnostic tool as well as a tool for the evaluation of tumor treatments.

For both of the above applications we approach the problem via the following steps:

- Construct reduced models of the multiple-compartment Bloch-Torrey partial differential equation (PDE) using homogenization methods.
- Invert the resulting reduced models for the biological parameters of interest: the cellular volume fraction in the first case, and the average distance between neighboring cells in the second case.

We obtained the following results.

- We generated fairly complicated meshes that can be used to simulate diffusion in cerebral gray matter. In the Finite Elements code, this required using the mesh generation software Salome, developed at the CEA Saclay. We are working on the problem of increasing the cellular volume fraction to a physically realistic level, which is difficult for the mesh generator because of the very small distances between the neurons.

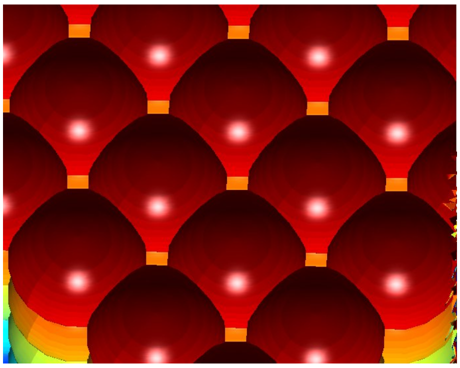


Figure 2. Computational domain for simulating tumor cells.

- We developed a homogenized model for the apparent diffusion coefficient (the slope of the log of the DMRI signal) of heterogenous cellular domains. An article on this topic has been submitted.
- We developed a reduce model of the complete DMRI signal (not just the slope as in the above) using more sophisticated homogenization methods. An article on this topic is under preparation.

DISCO Project-Team

6. New Results

6.1. Algorithmic study of linear functional systems

Participants: Alban Quadrat, Thomas Cluzeau [ENSIL, Univ. Limoges], Daniel Robertz [Univ. Aachen].

In [108], it is shown that every linear functional system (e.g., PD systems, differential time-delay systems, difference systems) is equivalent to a linear functional system defined by an upper block-triangular matrix of functional operators: each diagonal block is respectively formed by a generating set of the elements of the system satisfying a purely i-codimensional system. Hence, the system can be integrated in cascade by successively solving (inhomogeneous) i-codimensional linear functional systems to get a Monge parametrization of its solution space [110]. The results are based on an explicit construction of the grade/purity filtration of the module associated with the linear functional system. This new approach does not use involved spectral sequence arguments as is done in the literature of modern algebra [82], [83]. To our knowledge, the algorithm obtained in [34] is the most efficient algorithm existing in the literature of non-commutative algebra. It was implemented in the PURITYFILTRATION package developed in Maple (see Section 5.6) and in the `homalg` package of GAP 4 (see Section 5.7). Classes of overdetermined/underdetermined linear systems of partial differential equations which cannot be directly integrated by Maple can be solved using the PURITYFILTRATION package.

Given a linear multidimensional system (e.g., ordinary/partial differential systems, differential time-delay systems, difference systems), Serre's reduction aims at finding an equivalent linear multidimensional system which contains fewer equations and fewer unknowns. Finding Serre's reduction of a linear multidimensional system can generally simplify the study of structural properties and of different numerical analysis issues, and it can sometimes help solving the linear multidimensional system in closed form. In [13], Serre's reduction problem is studied for underdetermined linear systems of partial differential equations with either polynomial, formal power series or analytic coefficients and with holonomic adjoints in the sense of algebraic analysis [82], [83]. These linear partial differential systems are proved to be equivalent to a linear partial differential equation. In particular, an analytic linear ordinary differential system with at least one input is equivalent to a single ordinary differential equation. In the case of polynomial coefficients, we give an algorithm which computes the corresponding linear partial differential equation.

The connection between Serre's reduction and the decomposition problem [90], which aims at finding an equivalent linear functional system which is defined by a block diagonal matrix of functional operators, is algorithmically studied in [92].

In [111], algorithmic versions of Statford's results [114] (e.g., computation of unimodular elements, decomposition of modules, Serre's splitting-off theorem, Stafford's reduction, Bass' cancellation theorem, minimal number of generators) were obtained and implemented in the STAFFORD package. In particular, we show how a determined/overdetermined linear system of partial differential equations with either polynomial, rational, formal power series or locally convergent power series coefficients is equivalently to a linear system of partial differential in at most two unknowns. This result is a large generalization of the cyclic vector theorem which plays an important role in the theory of linear ordinary differential equations.

6.2. Boundary value problems for linear ordinary integro-differential equations

Participants: Alban Quadrat, Georg Regensburger.

In [61], we study algorithmic aspects of linear ordinary integro-differential operators with polynomial coefficients. Even though this algebra is not noetherian and has zero divisors, Bavula recently proved in [81] that it is coherent, which allows one to develop an algebraic systems theory. For an algorithmic approach to linear systems theory of integro-differential equations with boundary conditions, computing the kernel of matrices is a fundamental task. As a first step, we have to find annihilators, which is, in turn, related to polynomial solutions. We present an algorithmic approach for computing polynomial solutions and the index for a class of linear operators including integro-differential operators. A generating set for right annihilators can be constructed in terms of such polynomial solutions. For initial value problems, an involution of the algebra of integro-differential operators also allows us to compute left annihilators, which can be interpreted as compatibility conditions of integro-differential equations with boundary conditions. These results are implemented in MAPLE based on the `IntDiffOp` and `IntDiffOperations` packages. Finally, system-theoretic interpretations of these results are given and illustrated on integro-differential equations.

In [78], we develop linear algebra results needed for generalizing the composition of boundary problems to singular ones. We consider generalized inverses of linear operators and study the question when their product in reverse order is again a generalized inverse. This problem has been studied for various kinds of generalized inverses, especially for matrices. Motivated by our application to boundary problems, we use implicit representation of subspaces via "boundary conditions" from the dual space and this approach gives a new representation of the product of generalized inverses. Our results apply to arbitrary vector spaces and for Fredholm operators, the corresponding computations reduce to finite-dimensional problems, which is crucial for our implementation for boundary problem for linear ordinary differential equations.

In collaboration with Li Guo and Markus Rosenkranz [77], we study algebraic aspects of integro-differential algebras and their relation to so-called differential Rota-Baxter algebras. We generalize this concept to that of integro-differential algebras with weight. Based on free commutative Rota-Baxter algebras, we investigate the construction of free integro-differential algebras with weight generated by a regular differential algebra. The explicit construction is not only interesting from an algebraic point of view but is also an important step for algorithmic extensions of differential algebras to integro-differential algebras (compare with the related construction and the implementation of integro-differential polynomials in [72]). In this paper, we review also the construction of integro-differential operators, the algorithms for regular boundary problems and a prototype implementation in the Theorema system.

In [11], we adapt our factorization technique for boundary problems to study ruin probabilities and related quantities in renewal risk theory. The analysis is based on boundary problems for linear ordinary differential equations (on the half bounded interval from zero to infinity) with variable coefficients and the corresponding factorization of Green's operators. With this approach, we obtain closed-form and asymptotic expressions for discounted penalty functions under the more realistic assumption that the premium income depends on the present surplus of the insurance portfolio.

6.3. Symbolic methods for developing new domain decomposition algorithms

Participants: Thomas Cluzeau [ENSIL, Univ. Limoges], Victorita Dolean [Univ. Nice - Sophia-Antipolis], Frédéric Nataf [CNRS, Paris 6], Alban Quadrat.

Some algorithmic aspects of systems of partial differential equations based simulations can be better clarified by means of symbolic computation techniques. This is very important since numerical simulations heavily rely on solving systems of partial differential equations. For the large-scale problems we deal with in today's standard applications, it is necessary to rely on iterative Krylov methods that are scalable (i.e., weakly dependent on the number of degrees on freedom and number of subdomains) and have limited memory requirements. They are preconditioned by domain decomposition methods, incomplete factorizations and multigrid preconditioners. These techniques are well understood and efficient for scalar symmetric equations (e.g., Laplacian, biLaplacian) and to some extent for non-symmetric equations (e.g., convection-diffusion). But they have poor performances and lack robustness when used for symmetric systems of partial differential equations, and even more so for non-symmetric complex systems (fluid mechanics, porous media, ...). As a general rule, the study of iterative solvers for systems of partial differential equations as opposed to scalar partial differential equations is an underdeveloped subject. In [76], we aim at building new robust and efficient solvers, such as domain decomposition methods and preconditioners for some linear and well-known systems of partial differential equations based on algebraic techniques (e.g., Smith normal forms, Gröbner basis techniques, D-modules).

6.4. Noncommutative geometry approach to infinite-dimensional systems

Participant: Alban Quadrat.

In [105], [104], [103], it was shown how the fractional representation approach to analysis and synthesis problems developed by Vidyasagar, Desoer, Callier, Francis, Zames..., could be recast into a modern algebraic analysis approach based on module theory (e.g., fractional ideals, algebraic lattices) and the theory of Banach algebras. This new approach successfully solved open questions in the literature. Basing ourselves on this new approach, we explain in [107] why the non-commutative geometry developed by Alain Connes is a natural framework for the study of stabilizing problems of infinite-dimensional systems. Using the 1-dimensional quantized calculus developed in non-commutative geometry and results obtained in [105], [104], [103], we show that every stabilizable system and their stabilizing controllers naturally admit geometric structures such as connections, curvatures, Chern classes, ... These results developed in [59] are the first steps toward the use of the natural geometry of the stabilizable systems and their stabilizing controllers in the study of the important H_∞ and H_2-problems.

6.5. Stabilization of time-delay systems

Participants: Alban Quadrat, Arnaud Quadrat [SAGEM, MASSY].

In [60], we study the stabilization problem of a linear system formed by a simple integrator and a time-delay. We show that the stabilizing controllers of such a system can be be rewritten as the closed-loop system defined by the stabilizing controllers of the simple integrator and a distributed delay. This result is used to study tracking problems appearing in the study of inertially stabilized platforms for optical imaging systems.

6.6. Stabilization of MISO fractional systems with delays

Participants: Catherine Bonnet, Le Ha Vy Nguyen.

In order to yield the set of all the stabilizing controllers of a class of MISO fractional systems with delays by mean of Youla-Kucera parametrization regarding H_∞-stability, we are interested in determining coprime factorizations of the transfer function. Explicit expressions of left coprime factorizations and left Bézout factors are derived in [51]. On the other hand, right coprime factorizations exist, and we have obtained explicit expressions for several particular cases of the studied systems.

6.7. Stability analysis of (fractional) neutral systems with commensurate delays

Participants: Catherine Bonnet, Andre Fioravanti [UNICAMP], Le Ha Vy Nguyen.

Neutral time-delay systems may have chains of poles asymptotic to the imaginary axis. As the chains approach the axis, some systems are H_∞-unstable even though all the poles are in the left-half plane. For a class of such systems, H_∞-stability conditions were presented in [84]. While systems with no more than one chain of poles asymptotic to a set of points on the imaginary axis were exhaustedly studied, only a particular case of systems with multiple chains were considered. We continue the stability analysis for more general cases of the latter systems. Primary results on pole locations are obtained [53], [52]. Based on these results, H_∞-stability conditions have also been derived.

6.8. Matrix Norm Approach for Control of Linear Time-Delay Systems

Participants: Catherine Bonnet, André Fioravanti [UNICAMP], José Claudio Geromel [UNICAMP], Silviu Niculescu.

In [94], we have treated the time-delay linear systems control design in the framework of complete and partial information. We were able to find linear controllers that increase the first stability window imposing at the same time that the delay-free system is stable using some properties about the norms of the state-space matrices. Our method treated the design problem by numeric routines based on Linear Matrix Inequalities (LMI) arisen from classical linear time invariant system theory coupled together with a unidimensional search. Both the state and output feedback design, were solved. We have this year tried our method on a 'high-dimensional' example for which no existing direct method would be computationnally feasible.

6.9. Interval observer

Participants: Frederic Mazenc, Silviu Niculescu, Thach Ngoc Dinh, Olivier Bernard [Inria - Sophia-Antipolis], Eric Walter [CNRS - L2S - Supelec], Michel Kieffer [CNRS - L2S - Supelec].

We made several progresses in the domain of the construction of state estimators called interval observers.
1) We presented the design of families of interval observers for continuous-time linear systems with a point-wise delay after showing that classical interval observers for systems without delays are not robust with respect to the presence of delays and that, in general, for linear systems with delay, the classical interval observers endowed with a point-wise delay are unstable. We proposed a new type of design of interval observers enabling to circumvent these obstacles. It incorporates distributed delay terms [26].

2) We considered a family of continuous-time systems that can be transformed through a change of coordinates into triangular systems. By extensively using this property, we constructed interval observers for nonlinear systems which are not cooperative and not globally Lipschitz. For a narrower family of systems, the interval observers possess the Input to State Stability property with respect to the bounds of the uncertainties [42], [21].

3) For the first time, we addressed in [44] the problem of constructing interval observers for discrete-time systems. Under a strong assumption, we proposed time-invariant interval observers for a very broad family of systems. In a second step, we have shown that, for any time-invariant exponentially stable discrete-time linear system with additive disturbances, time-varying exponentially stable discrete-time interval observers can be constructed. The latter result relies on the design of time-varying changes of coordinates which transform a linear system into a nonnegative one.

4) We considered continuous-time linear systems with additive disturbances and discrete-time measurements. First, we constructed a standard observer, which converges to the state trajectory of the linear system when the maximum time interval between two consecutive measurements is sufficiently small and there are no disturbances. Second, we constructed interval observers allowing to determine, for any solution, a set that is guaranteed to contain the actual state of the system when bounded disturbances are present [46].

6.10. New reduction model approach

Participants: Frederic Mazenc, Silviu Niculescu, Mounir Bekaik, Dorothee Normand-Cyrot [CNRS - L2S - Supelec], Claudio de Persis [Sapienza University of Rome], Miroslav Krstic [Univ. of California].

We considered several distinct problems entailing to the reduction model approach. Let us recall that this technique makes it possible to stabilize systems with arbitrarily large pointwise or distributed delay.

1) We proposed a new construction of exponentially stabilizing sampled feedbacks for continuous-time linear time-invariant systems with an arbitrarily large constant pointwise delay in the inputs. Stability is guaranteed under an assumption on the size of the largest sampling interval. The proposed design is based on an adaptation of the reduction model approach. The stability of the closed loop systems is proved through a Lyapunov-Krasovskii functional of a new type, from which is derived a robustness result [28], [50].

2) For linear systems with pointwise or distributed delays in the inputs which are stabilized through the reduction approach, we proposed a new technique of construction of Lyapunov-Krasovskii functionals. These functionals allow us to establish the ISS property of the closed-loop systems relative to additive disturbances [27], [49].

3) We proposed a solution to the problem of stabilizing nonlinear systems with input with a constant pointwise delay and state-dependent sampling. It relies on a recursive construction of the sampling instants and on a recent variant of the classical reduction model approach. The state feedbacks that are obtained do not incorporate distributed terms [43].

6.11. Analysis of neutral systems

Participants: Frederic Mazenc, Hiroshi Ito [Kyushu Institute of Technology].

1) For nonlinear systems with delay of neutral type, we developped a new technique of stability and robustness analysis. It relies on the construction of functionals which make it possible to establish estimates of the solutions different from, but very similar to, estimates of ISS or iISS type. These functionals are themselves different from, but very similar to, ISS or iISS Lyapunov-Krasovskii functionals. The approach applies to systems which do not have a globally Lipschitz vector field and are not necessarily locally exponentially stable. We apply this technique to carry out a backstepping design of stabilizing control laws for a family of neutral nonlinear systems [22], [45].

2) We extended the previous result to the problem of deriving the iISS property for dynamical networks with neutral, retarded and communication delay [41].

6.12. Hyperbolic systems

Participants: Frederic Mazenc, Christophe Prieur [GIPSA-Lab CNRS].

We considered a family of time-varying hyperbolic systems of balance laws. The partial differential equations of this family can be stabilized by selecting suitable boundary conditions. For the stabilized systems, the classical technique of construction of Lyapunov functions provides a function whose derivative along the trajectories of the systems may be not negative definite. In order to obtain a Lyapunov function with a negative definite derivative along the trajectories, we transform this function through a so-called "strictification" approach, which gives a time-varying strict Lyapunov function. It allows us to establish asymptotic stability in the general case and a robustness property with respect to additive disturbances of Input-to-State Stability type [32].

6.13. Time-varying systems with delay

Participants: Frederic Mazenc, Silviu Niculescu, Mounir Bekaik, Michael Malisoff [Departement of Mathematics - LSU].

1) We solved aproblem of state feedback stabilization of time-varying feedforward systems with a pointwise delay in the input. The approach relies on a time-varying change of coordinates and Lyapunov-Krasovskii functionals. The result applies for any given constant delay, and provides uniformly globally asymptotically stabilizing controllers of arbitrarily small amplitude. The closed-loop systems enjoy Input-to-State Stability properties with respect to additive uncertainty on the controllers. The work is illustrated through a tracking problem for a model for high level formation flight of unmanned air vehicles [48], [24].

2) We addressed the problem of stabilizing systems belonging to a family of time-varying nonlinear systems with distributed input delay through state feedbacks without retarded term. The approach we adopted is based on a new technique that is inspired by the reduction model technique. The control laws we obtained are nonlinear and time-varying. They globally uniformly exponentially stabilize the origin of the considered system. We illustrate the construction with a networked control system [25].

6.14. Positive invariance for time delay systems

Participants: Sorin Olaru [correspondent], Silviu Niculescu [CNRS (LSS)], Georges Bitsoris [University of Patras, Greece].

A new concept of positive invariance has been established in the original state space for discrete time dynamical systems. Furthermore, the necessary and sufficient algebraic condition for such properties have been derived allowing a direct test using basic linear programming arguments. In a recent work, the rigid positive invariance has been relaxed toward a cyclic invariant concept [18].

6.15. Predictive control for networked control systems

Participants: Sorin Olaru [correspondent], Silviu Niculescu [CNRS (LSS)], Warody Lombardi [INSA Lyon].

The work on the networked control system modeling lead to the establishement of a solid framework based on linear difference inclusion. Subsequently via set invariance and optimization based techniques, a design procedure has been proposed to deal with the real time constrained feedback control. Is worth to be mentioned that the robust feasibility and control performances are enforced via inverse optimality principles [19].

6.16. Reduced order H_∞-controllers synthesis with explicit constraints handling

Participants: Guillaume Sandou [correspondent], Gilles Duc [Suplec (E3S), Control Department], Mohamed Yagoubi [Ecole des Mines de Nantes].

Efficient dedicated methods have been developed for Hinfinity controller synthesis. However, such methods require translating the design objectives using weighting filters, whose tuning is not easy; in addition they lead to high order controllers which have to be reduced. Previous works have dealt with these two problems separately with the help of Particle Swarm Optimization: optimization of filter tunings for a full order synthesis and reduced order synthesis with fixed filters. In recent works, we have considered the solution to both problems in one shot. The constraints of the problem are explicitly taken into account in the synthesis problem, thanks to the use of Particle swarm optimization which does not require any specific expression for costs and constraints [63].

6.17. Robust optimization for energy management

Participants: Guillaume Sandou [correspondent], Philippe Dessante [Suplec (E3S), Energy Department], Marc Petit [Suplec (E3S), Energy Department].

The optimization of energy networks and the solution to Unit Commitment problems are one of the main collaborations between the Control and Energy Departments of Supelec. Robust optimization has been used to take into account the uncertainties which are observed on the consumer demand, the cost function, and the maximum capacity [66], [73].

6.18. Firefly optimization for the synthesis of controllers and the identification of systems

Participants: Guillaume Sandou [correspondent], Alfonso Goches Sanchez [Suplec (E3S), Control Department].

Firefly optimization is a new optimization algorithm which has appeared in 2009. This algorithm belongs to the class of metaheuristic algorithms. As such algorithms can optimized any cost and functions, firefly optimization has been tested for the optimization of PID controllers (with no reformulations of specifications) and the identification of nonlinear systems.

6.19. Receding horizon based controllers for the energy management in complex systems

Participants: Guillaume Sandou [correspondent], Sorin Olaru, Silviu Niculescu, Emmanuel Witrant [Gips-Lab, Grenoble].

The use of receding horizon based controllers is a good trend to extend the optimization results of a complex system in a closed loop framework. To prove the viability and the efficiency of the approach, several real life examples have been tested. Among them are the district heating networks and the mining ventilation system.

6.20. Particle Swarm Optimization for the optimization of feasibility domain volumes

Participants: Guillaume Sandou [correspondent], Mohamad-Taki Asghar [Suplec (E3S), Control Department].

It is a well-known fact that using mu-analysis for the computation of a guaranteed stability domain gives the largest hyper-rectangle included in the real stability domain (which is impossible to compute). However, the results strongly depend on the choice which has been made for the nominal system and the parameterization of the uncertainties. In this study, these choices are considered as optimization variables. The goal is now to find the best parameterization of the problem to get the largest stability domain. The optimization has been done using Particle Swarm Optimization.

6.21. Model of reaction networks

Participants: Georg Regensburger, Stefan Müller [RICAM, Linz].

In [100], we propose a notion of generalized mass action systems that could serve as a more realistic model for reaction networks in intracellular environments; classical mass action systems capture chemical reaction networks in homogeneous and dilute solutions. We show that several results of chemical reaction network theory carry over to the case of generalized mass action kinetics. Our main result gives conditions for the existence of a unique positive steady state for arbitrary initial conditions and independent of rate constants in this generalized setting. The conditions are formulated in terms of sign vectors (oriented matroids) of the stoichiometric and kinetic-order subspace and face lattices of related cones. We also give necessary and sufficient conditions for multistationarity, which is an important property in many applications, for example, in connection with cell differentiation.

6.22. Control of aircraft dynamics

Participants: Frederic Mazenc, Michael Malisoff [Departement of Mathematics - LSU], Aleksandra Gruszka [Departement of Mathematics - LSU].

We have worked on several models describing physical devices.

1) We studied a kinematic model that is suitable for control design for high level formation flight of UAVs [16], [40]. We designed controllers that give robust global tracking for a wide class of reference trajectories in the sense of input-to-state stability while satisfying amplitude and rate constraints on the inputs.

2) We studied feedback tracking problems for the planar vertical takeoff and landing (PVTOL) aircraft dynamics, which is a benchmark model in aerospace engineering. We provided a survey of the literature on the model. Then we constructed new feedback stabilizers for the PVTOL tracking dynamics. The novelty of our work is in the boundedness of our feedback controllers and their applicability to cases where the velocity measurements may not be available, coupled with the uniform global asymptotic stability and uniform local exponential stability of the closed loop tracking dynamics, and the input-to-state stable performance of the closed loop tracking dynamics with respect to actuator errors [15].

3) We solved a stabilization problem for an important class of feedback controllers that arise in curve tracking problems for robotics. Previous experimental results suggested the robust performance of the control laws under perturbations. Consequently, we used input-to-state stability to prove predictable tolerance and safety bounds that ensure robust performance under perturbations and time delays. Our proofs are based on an invariant polygon argument and a new strict Lyapunov function design [20].

6.23. Study of chemostat models

Participants: Frederic Mazenc, Michael Malisoff [Departement of Mathematics - LSU].

We provided a study of chemostat models in which two or more species compete for two or more limiting nutrients. First we considered the case where the nutrient flow and species removal rates and input nutrient concentrations are all given positive constants. In that case, we used Brouwer fixed point theory to give conditions guaranteeing that the models admit globally asymptotically stable componentwise positive equilibrium points. For cases where the dilution rate and input nutrient concentrations can be selected as controls, we used Lyapunov methods to prove that many different possible componentwise positive equilibria can be made globally asymptotically stable. We demonstrated our methods in simulations [23].

6.24. Modeling and control of Acute Myeloid Leukemia

Participants: José Luis Avila Alonso, Annabelle Ballesta [BANG project-team], Frédéric Bonnans [COMMANDS project-team], Catherine Bonnet, Jean Clairambault [BANG project-team], Xavier Dupuis [COMMANDS project-team], Pierre Hirsch [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Jean-Pierre Marie [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Faten Merhi [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Silviu Niculescu, Hitay Özbay [Bilkent University, Ankara, Turkey], Ruoping Tang [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris].

We have continued this year our work on modeling healthy and pathological hematopoiesis [36]. A. Ballesta has performed some experiments on patient fresh cell cultures in order to identify parameters of our model of acute myeloblastic leukemia (AML). To evaluate therapies, she also considered patient fresh cell cultures under anticancer drugs.

DOLPHIN Project-Team

6. New Results

6.1. On Optimizing a Bi-objective Flowshop Scheduling Problem in Uncertain Environment

Participants: Arnaud Liefooghe, Laetitia Jourdan, El-Ghazali Talbi

Existing models from scheduling often over-simplify the problems appearing in real-world industrial situations. The original application is often reduced to a single-objective one, where the presence of uncertainty is neglected. In [23], we focus on multi-objective optimization in uncertain environments. A bi-objective flowshop scheduling problem with uncertain processing times is considered. An indicator-based evolutionary algorithm is proposed to handle these two difficulties (multiple objectives and uncertain environment) at the same time. Four different strategies, based on uncertainty-handling quality indicators, are proposed in the paper. Computational experiments are performed on a large set of instances by considering different scenarios with respect to uncertainty. We show that an uncertainty-handling strategy is a key issue to obtain good-quality solutions, and that the algorithm performance is strongly related to the level of uncertainty over the environmental parameters.

6.2. New Price settings models in the energy field

Participants: L. Brotcorne, S. Afsar

The electricity supply industry is facing in many countries a restructuring process towards deregulation and competition. In that context classical marginal cost based approaches based on estimation of cost production function and demand functions are not well-suited anymore. Indeed, the energy prices have to be defined not only to retrieve the production costs but also in order to take into account the consumer behavior. Consumers make their choice of service, or of energy provider in order to minimize their disutility values. Failing to recognize that may lead to tremendous lack on revenues. In order to capture this hierarchical decision process where a leader (the energy provider) takes explicitly into account the reaction of a follower (the consumers) in his decision process. The energy pricing problems addressed in this are modeled as bilevel programs.

6.3. Bi-level formulation for a Long-Distance Freight Transportation Problem

Participants: M. Diaby, L. Brotcorne, E.-G. Talbi.

A company wants to convey different types of products from origin i to points of destination j. It can deliver the goods itself or hire a transport company, and subcontract part of the application. The transport company must offer attractive prices while aiming to maximize its profit. The aim of this problem is to determine rates that allow the carrier to maximize its revenues and remain affordable for the customer. The problem is modeled as a bilevel program at the first level, the carrier (leader) wants to maximize its revenues; at the second level, the client. An exact and an evolutionary solution approaches are developped.

6.4. On Local Search for Bi-objective Knapsack Problems

Participant: A. Liefooghe.

In [26], a local search approach is proposed for three variants of the bi-objective binary knapsack problem, with the aim of maximizing the total profit and minimizing the total weight. First, an experimental study on a given structural property of connectedness of the efficient set is conducted. Based on this property, a local search algorithm is proposed and its performance is compared against exact algorithms in terms of running time and quality metrics. The experimental results indicate that this simple local search algorithm is able to find a representative set of optimal solutions in most of the cases, and in much less time than exact algorithms.

6.5. Convergent methods based on aggregation in mathematical models

Participant : François Clautiaux

We designed several algorithms to aggregate variables in integer linear programs. Our methods first solve aggregated models, and converge to the optimal solution of the initial problem by iteratively refining the model.

The first method applies on a large network flow models that use a pseudo-polynomial number of variables. It is based on an initial aggregation of the vertices of the model and its iterative refinement using different optimization techniques. This led to dramatical improvements for a special case of vehicle routing problem. We proposed several theoretical results regarding convergence, suitable discretizations, wort-case analysis and approximation algorithms [44].

The second method applies on column generation approaches for the cutting-stock problem. Our algorithm links groups of dual variables by linear constraints, leading to a problem of smaller dimension, whose solutions are dual-feasible for the initial problem. The corresponding "inner approximation" is iteratively refined by splitting the groups into smaller groups until an optimal dual solution is found. This method allows to produce a valid lower bound at each iteration, which is not the case for classical column-generation schemes [58].

6.6. Investigating the Optimization Goal of Indicator-Based Multiobjective Search

Participant from DOLPHIN: Dimo Brockhoff; External Participants: Heike Trautmann and Tobias Wagner (TU Dortmund University, Germany)

Using a quality indicator in the environmental selection step of evolutionary multiobjective optimization (EMO) algorithms to indicate which solutions shall be kept in the algorithms' population and which should be deleted, introduces a certain search bias. Instead of an “arbitrary” subset of the Pareto front, such (quality) indicator based search algorithms aim at approximating the set of μ solutions that optimizes a given indicator, for which the term *optimal μ-distribution* has been introduced [63]. Also for performance assessment with respect to a given indicator, knowledge about the optimal μ-distributions is helpful as interpreting the *achieved* indicator values with respect to the best *achievable* value becomes possible. For the hypervolume indicator, several results on these optimal μ-distributions are known [63], [62], [75], [69], [70], [61] [64], but the understanding of the optimization goal for other indicators is less developed. Recently, we started to investigate the optimal μ-distributions, both theoretically and numerically, for the so-called $R2$ indicator [79]—another often recommended quality indicator [90]. Instead of the binary version of [79] that takes two solution sets and assigns them a certain quality, we thereby investigated an equivalent unary indicator where one (reference) set is always fixed.

First experiments for problems with two objectives and connected Pareto fronts have been presented in [37] which won the best paper award within the EMO track at GECCO'2013 [5]. Further investigations on problems with disconnected Pareto fronts have been submitted to the Evolutionary Computation journal [72]. We also studied in more detail how the parameters of the $R2$ indicator such as the ideal point or the distribution of weight vectors can be used to change the optimization goal [86] and correspondingly proposed the algorithm $R2$-EMOA which is able to steer the search towards preferred regions of the Pareto front by optimizing the $R2$ indicator directly in its environmental selection [85], [72].

6.7. Runtime Analyses of Interactive Evolutionary Algorithms

Participant from DOLPHIN: Dimo Brockhoff; External Participants: Manuel López-Ibáñez (Université Libre de Bruxelles, Belgium), Boris Naujoks (Cologne University of Applied Sciences, Germany), and Gunter Rudolph (TU Dortmund University, Germany)

[5] See http://www.sigevo.org/gecco-2012/papers.html.

If a decision maker (DM) expresses preferences, e.g., towards certain points or regions of the search space, during the algorithm run, we call such an algorithm *interactive*. Interactive algorithms are frequently used in the field of multi-criteria decision making, but theoretical results on interactive evolutionary multiobjective algorithms (EMOAs) have not been derived until recently. In [36], we started to analyze interactive versions of an evolutionary algorithm with plus-selection and a population size of one, the so-called iRLS and $i(1+1)$EA. On two pseudo-boolean problems, recently used for theoretically analyzing EMOAs, we could prove upper bounds on the expected runtime of the two mentioned algorithms and on the number of times, the DM is asked about his/her preferences until the most-preferred search point is found. The analyzes showed that the internal value function of the DM has a strong, non-desired influence on the algorithms' runtimes and that the number of questions to the DM are too high for a practical relevant algorithm. It is an open question which algorithm designs are necessary to circumvent these two drawbacks.

6.8. Benchmarking of CMA-ES Variants for Numerical Blackbox Optimization

Participant from DOLPHIN: Dimo Brockhoff; External Participants: Anne Auger and Nikolaus Hansen (Inria Saclay - Ile-de-France)

The covariance matrix adaptation evolution strategy (CMA-ES) is one of the state-of-the-art optimization algorithms for numerical single-objective blackbox optimization [81], [80] [67]. Previously, we proposed to use so-called mirrored mutations to generate new candidate solutions in evolution strategies which turned out to increase the convergence rate for certain variants [71], [65], [66]. Another recent approach to speed up the CMA-ES is to perform an active (i.e. negative) covariance matrix update [60]. In [32], [35], [34], [33], we tested empirically how the combination of mirrored mutations and active CMA-ES perform on the COCO framework [77], [78]. It turned out that both concepts complement each other well without a significant decrease in performance on any of the 24 test functions. Moreover, the main improvement over the standard CMA-ES could be shown to come from the active covariance matrix adaptation while the addition of mirrored mutations only slightly improves the algorithm.

6.9. Self-adaptive method for a three-objective vector-packing problem

Participants: Nadia Dahmani, François Clautiaux, El-Ghazali Talbi

We introduced a new multi-objective packing problem (MOBPP), in which we optimize the number of bins, the maximum weight of a bin, and the loading balancing. We studied the impact and the combination of two complementary decoding strategies for this problem. A feature of our work is to insert the parameters of the decoders in the representation of the solution. It leads to self-adaptive meta-heuristics, where the algorithm iteratively adapts the parameters during the search. We embedded our approaches in a local search and an evolutionary algorithm for the MOBPP. A comprehensive set of experiments were performed on various benchmarks inspired from the literature. Results confirm that our methods lead to more effective muti-objective metaheuristics for this problem.

6.10. Multiple Neighborhood Exploration Through Adaptive Search

Participants: Bilel Derbel, Houda Derbel, El-Ghazali Talbi, Hiba Yahyaoui

Variable neighborhood descent (VND) and its several variants are based on the systemic change of neighborhoods within the search. It is well known that the performance of a VND-like algorithm highly depends on the order/way the neighborhoods are alternated. In this work, we focus on designing new meta-strategies for deciding what neighborhood structure to apply through the search. Two new approaches are proposed to tackle this issue. In the first approach [41], we model the search by considering the neighborhood tree induced by alternating the use of different structures within a local search descent. We investigate the issue of designing a search strategy operating at the neighborhood tree level by exploring different paths of the tree in a heuristic way. We show that allowing the search to 'backtrack' to a previously visited solution and resuming the iterative variable neighborhood descent by 'pruning' the already explored neighborhood branches leads to the design of effective and efficient search heuristics. In the second approach, we investigate deterministic and randomized adaptive strategies for selecting the next neighborhood to apply at runtime. The adaptive strategies are based on computing a reward for each neighborhood with respect to the observed average ratio of solution quality and time cost.

6.11. CoBRA: A cooperative coevolutionary algorithm for bi-level optimization

Participants: François Legillon, Arnaud Liefooghe, El-Ghazali Talbi

In [43] we present CoBRA, a new evolutionary algorithm, based on a coevolutionary scheme, to solve bi-level optimization problems. It handles population-based algorithms on each level, each one cooperating with the other to provide solutions for the overall problem. Moreover, in order to evaluate the relevance of CoBRA against more classical approaches,a new performance assessment methodology, based on rationality, is introduced. An experimental analysis is conducted on a bi-level distribution planning problem, where multiple manufacturing plants deliver items to depots, and where a distribution company controls several depots and distributes items from depots to retailers. The experimental results reveal significant enhancements, particularly over the lower level, with respect to a more classical approach based on a hierarchical scheme.

6.12. Neutrality in the Graph Coloring Problem

Participants: Marie-Eléonore Marmion, Aymeric Blot, Laetitia Jourdan, and Clarisse Dhaenens

The graph coloring problem is often investigated in the literature. Many insights about the existence of many neighboring solutions with the same fitness value are raised but as far as we know, no deep analysis of this neutrality has ever been conducted in the literature. We have quantified the neutrality of some hard instances of the graph coloring problem. This neutrality property has to be detected as it impacts the search process. Indeed, local optima may belong to plateaus that represents a barrier for local search methods. We also aim at pointing out the interest of exploiting neutrality during the search. Therefore, a generic local search dedicated to neutral problems (NILS) and previously tested on flowshop problems, is performed and tested on several hard instances. Results show that taking into account neutrality allows to obtain better results than when not considering it.

6.13. Local Search in the Context of Classification Rule Mining

Participants: Julie Jacques, Laetitia Jourdan, Clarisse Dhaenens

Many multi-objective algorithms have been proposed to solve the classification rule mining problem; the vast majority of them are based on genetic algorithms. We propose an algorithm, MOCA - Multi-Objective Classification Algorithm -, to solve this problem. The originality of MOCA is to be a dominance-based multi-objective local search (DMLS) using a Pittsburgh representation of rules. We evaluated several DMLS implementations and neighborhood operators on literature datasets and one real dataset. Then we compared the best obtained algorithm against several efficient approaches of the literature. The experiments show that the proposed approach is very competitive in comparison to other algorithms tested. Moreover, our approach is able to deal with very large real datasets and manages to have a good accuracy.

6.14. MOCA-I: discovering rules and guiding decision maker in the context of partial classification in large and imbalanced datasets

Participants: Julie Jacques, Laetitia Jourdan, Clarisse Dhaenens

In this work we focus on the modeling and the implementation as a multi-objective optimization problem of a Pittsburgh classification rule mining algorithm adapted to large and imbalanced datasets, as encountered in hospital data. Indeed hospital data comes with problems such as class imbalance, volumetry or inconsistency, and optimization approaches have to take into account such specificities. We present MOCA-I, an adaptation of MOCA dedicated to this kind of problems. We propose its implementation as a dominance-based local search in opposition to existing multi-objective approaches based on genetic algorithms. We associate to this algorithm an original post-processing method based on the ROC curve to help the decision maker to choose the most interesting set of rules. Our approach is currently compared to state-of-the-art classification rule mining algorithms (both classic approaches such as C4.5 and optimization approaches), giving as good or better preliminary results, using less parameters. Moreover, our approach has been compared to C4.5 and C4.5-CS on a real dataset (hospital data) with a larger set of attributes, giving the best results. The complete evaluation is still going on.

6.15. A method to combine combinatorial optimization and statistics to mine high-throughput genotyping data

Participants : Julie Hamon, Clarisse Dhaenens, Julien Jacques (MODAL)

In the context of genomic analysis (collaboration with Genes Diffusion), dealing with high-throughput genotyping data, the objective of our study is to select a subset of SNPs (single nucleotide polymorphisms) explaining a trait of interest. We propose a method combining combinatorial optimization and statistics to extract a subset of interesting SNPs. The combinatorial part aims at exploring in an efficient way the large search space induced by the large number of possible subsets and statistics are used to evaluate the selection. We propose a first method based on an ILS (iterated local search) and using a regression. Three criteria used to evaluate the quality of the regression are compared. One of them (the k-fold validation) shows better performance. We also compare this approach to classical statistical approaches on simulated datasets. Results are promising as the proposed approach outperforms most of these statistical approaches [51].

6.16. Design and implementation of performance or energy-aware parallel optimization algorithms

Problems in practice are nowadays becoming more and more complex and time-intensive and their resolution requires to harness more and more computational resources. In parallel, the recent advances in hardware architecture enable to provide such required tremendous computational power through massively multi-core and GPU infrastructures. Such huge amount of cores is often provided through heterogeneous single or multi-clusters. The exploitation of such infrastructures clearly poses two fundamental and conflictual issues which are two major challenging perspectives of the Dolphin project that have been investigated during the 2012 year: (1) *Performance-aware issue*: how to design, implement and validate efficient and effective algorithms for such target machines to solve large size combinatorial optimization problems? ; (2) *Energy-aware issue*: using a large amount of computational resources for the deployment of large scale algorithms is energy-consuming. Therefore, the second issue is how to deal with the performance issue with a minimized cost in terms of energy consumption?
To deal with these issues, we have proposed new approaches summarized in the following sections.

6.16.1. Design and implementation of performance-aware optimization algorithms

In order to allow one to solve large size combinatorial optimization problems, we have revisited the design and implementation of meta-heuristics and exat (B&B) algorithms for two major hardware platforms: heterogeneous multi and many-core clusters and computational grids including multiple clusters.

- **Multi-core GPU-based hybrid meta-heuristics** - *Participants: T-V. Luong, N. Melab and E-G. Talbi.*

In [28], we have revisited the design and implementation of respectively single-solution and population-based meta-heuristics for single-core CPU coupled with a GPU device. We have investigated and proposed a new guidline for combining multi-core and GPU computing for hybrid meta-heuristics. Efficient approaches have been proposed for CPU-GPU data transfer optimization and task repartition between the GPU device and the CPU cores. Extensive experiments have been performed on an 8-core CPU coupled with a GPU card using Ant colonnies combined with a local serach applied to the Quadratic Assignment Problem (QAP). The reported results show that the use of multi-core computing, in addition to GPU computing, provides a performance improvement of up to 75%.

- **GPU-accelerated Branch-and-Bound algorithms** - *Participants: I. Chakroun and N. Melab.*
  Branch-and-Bound (B&B) algorithms are based on an implicit enumeration of a dynamically built tree-based search space. Pruning tree nodes (sub-problems) is traditionally used as a powerful mechanism to reduce the size of the explored search space. Such mechanism requires to perform the bounding operation which consists in applying a lower bound function to the generated sub-problems. Preliminary experiments performed on the Flow-Shop scheduling problem (FSP) have shown that the bounding operation consumes over 98% of the execution time of the B&B algorithm. Therefore, we have investigated the use of GPU computing as a major complementary way to speed up the search. We have revisited the design and implementation of the parallel bounding model for FSP on GPU accelerators dealing with two major issues: (1) thread divergence caused by the highly irregular nature of the explored tree and the SIMD execution model of GPU ; (2) data access optimization required for mapping efficiently different data structures on the hierarchy of memories provided in the GPU device. In [14], we have proposed a GPU-based parallel bounding model together with a data refactoring approach to deal with thread divergence. In [45] (an extended version submitted to the CCPE journal is being revised), we have proposed an efficient data optimization strategy based on a deep analysis of the complexity of the different data structures of the FSP lower bound algorithm in terms of memory size and access latency. The different proposed approaches for the two issues have been extensively experimented using and Nvidia Tesla C2050 GPU card. Compared to a CPU-based execution, accelerations up to more than $\times 100$ are achieved for large problem instances.

- **Peer-to-peer Branch-and-Bound algorithms** - *Participants: T-T. Vu, B. Derbel and N. Melab.*
  To deal with dynamic load balancing in large scale distributed systems, we have proposed in [50] to organize computing resources following a logical peer-to-peer overlay and to distribute the load according to the so-defined overlay. We have used a tree as a logical structure connecting distributed nodes and we balance the load according to the size of induced subtrees. We have conducted extensive experiments involving up to 1000 computing cores and provided a throughout analysis of different properties of our generic approach for two different applications, namely, the standard Unbalanced Tree Search and the more challenging parallel Branch-and-Bound algorithm. Substantial improvements are reported in comparison with the classical random work stealing and two finely tuned application specific strategies taken from the literature.

6.16.2. Design and implementation of energy-aware optimization algorithms

- Participants: Y. Kessaci, N. Melab and E-G. Talbi.

Cloud computing is an emerging computer science paradigm of distributed computing in which applications, data and infrastructures are proposed as a service that can be consumed in a ubiquitous, flexible and transparent way. Cloud computing brings with it such benefits via cloud managers which hide to the user some complex and challenging issues such as scheduling. However, the solutions to these issues provided in cloud managers are sometimes limited. For instance, the scheduling approach proposed in many cloud managers like OpenNebula is limited regarding the criteria taken into account. Energy consumption, which is highly critical for many applications such as High Performance Computing (HPC), is rarely considered. In [42] (selected for a special issue of FGCS journal) , we have addressed energy-aware scheduling of energy and time-consuming applications for cloud infrastructures. We have proposed a multi-start parallel local search heuristic for cloud managers (EMLS-ONC) with the focus put on OpenNebula. EMLS-ONC has been experimented using different (VMs) arrival scenarii and different hardware infrastructures. The results show that EMLS-ONC outperforms the scheduler provided in OpenNebula by a significant margin in terms of energy consumption and number of scheduled VMs.

GAMMA3 Project-Team

3. New Results

3.1. Validité des éléments finis usuels

Participants: Houman Borouchaki, Paul-Louis George [correspondant].

éléments finis-éléments finis généralisés-P1-P2-Q1-Q2-Bézier

On continue l'étude sur les conditions assurant la validité géométrique des éléments finis usuels de degré 1 et 2. La formulation éléments finis ne conduisant pas toujours à une conclusion simple, on formule les éléments finis sous leur forme de Bézier. Ceci conduit à exhiber des conditions suffisantes (parfois nécessaires et suffisantes) de validité des éléments, c'est-à-dire des conditions garantissant la positiivité de leur jacobien. Pour les éléments de degré 2, on donne l'interprétion géométrique de ces conditions. Les éléments étudiés sont le triangle à 3 nœuds, le triangle à 6 nœuds, le quadrilatère à 4 nœuds et les quadrilatère à 8 et 9 nœuds, le tétraèdre à 4 nœuds et le tétraèdre à 10 nœuds puis les pentaèdres à 6, 15 et 18 nœuds et les hexaèdres à 8, 27 et 20 nœuds.

On regarde ensuite les éléments finis généralisés déduits d'une formulation en Bézier rationnels puis basés sur des fonctions B-splines et Nurbs.

3.2. Maillages tétraédriques de grande taille

Participants: Houman Borouchaki, Paul-Louis George [correspondant], Loïc Maréchal.

Triangulation-tétraèdre p1-Hilbert- Maillage de grande taille

Le comportement en complexité des algorithmes de triangulation sur les "gros" maillage nous amène à utiliser les algorithmes de renumérotation de type Hilbert qui minimisent les défauts de cache. Cette technique est également utilisée comme aide à l'optimisation des "gros" maillages avec des gains en temps important. L'algorithme de renumérotation est multi-cœurs.

Des triangulations de plusieurs dizaines de millions de sommets sont construites en utilisant un "simple" ordinateur. La vitesse d'insertion frole le million de tétraèdres à la seconde.

Par coquetterie (et pour améliorer la robustesse dans l'absolu), on regarde ce que donne nos méthodes quand on construit des maillages de plus de un milliard de tétraèdres en séquentiel (une machine de un Tera de mémoire est utilisée). On vérifie que la taille des cavités peut être arbitrairement grande ce qui nécessite une programmation plus délicate permettant de traiter ces cas peu courants dans les situations habituelles.

3.3. Surface meshing with metric gradation control

Participants: Patrick Laug [correspondant], Houman Borouchaki.

Scientific computing requires the automatic generation of high quality meshes, in particular isotropic or anisotropic meshes of surfaces defined by a CAD modeler. For this purpose, two major approaches are called direct and indirect. Direct methods (octree, advancing-front or paving) work directly in the tridimensional space, while indirect methods consist in meshing each parametric domain and mapping the resulting mesh onto the composite surface. Using the latter approach, we propose a general scheme for generating "geometric" (or geometry-preserving) meshes by means of metrics. In addition, we introduce a new methodology for controlling the metric gradation in order to improve the shape quality. Application examples have shown the capabilities of this approach.

3.4. Metric field interpolation

Participants: Patrick Laug [correspondant], Houman Borouchaki.

To solve a physical problem formulated in terms of partial differential equations, the finite element method is generally used, based on a spatial discretization, or *mesh*, of the domain studied. Local adaptations of meshes to the behavior of the physical phenomena can improve the accuracy to the computed solutions, and in particular it is possible to capture high variations of the solution in specific areas while maintaining a reasonable number of degrees of freedom. In an initial phase, a mesh of the domain is built by using any particular method, then a first calculation of the solution of the problem is made. After choosing an appropriate criterion (Hessian and/or gradient of the solution, error estimate in general), areas that must be adapted by refinement or coarsening are detected in the initial mesh, and a new mesh is generated which is better adapted to the problem. This process is iterated until obtaining a mesh which satisfies the specified criterion (for which the finite element error is bounded by a specified threshold).

In practice, via an *a posteriori* analysis of the finite element error, a discrete map of sizes or metrics is set to the mesh vertices. This discrete size or metric field is made continuous by interpolating on the mesh, and the new mesh is generated according to this new field. In general, for a given point of the domain, a mesh element containing this point is found, and the interpolation of the size or metric field at this point is made from the sizes or metrics associated with the vertices of the containing element. For a scalar size field, the interpolation is straightforward by considering any interpolation scheme (for instance linear or geometric). On the other hand, the same scheme cannot be applied in the case of metrics representing a tensor field. However, several approaches have been proposed based on the link between a size and the corresponding metric and, in most cases, the interpolation scheme for sizes is applied to a power or the logarithm of the metrics. In particular, as a size h is represented by the isotropic metric $\mathcal{M} = \frac{1}{h^2}\,\mathcal{I}$, where $\mathcal{I}$ is the identity matrix, a possible link consists in approximating the size by $\mathcal{M}^{-\frac{1}{2}}$, then applying the size interpolation scheme to this new metric and finally recovering the interpolated metric. These schemes are still an approximation and require the calculation of the eigenvalues of $\mathcal{M}$ which is generally costly.

In this work, a new method for interpolating discrete metric fields is proposed. It is based on the "natural decomposition" of metrics using the LU factorization. With this decomposition, for each metric, the natural sizes along particular (or natural) directions can be retrieved, thus the size interpolation scheme can be applied to both natural directions and sizes, and the interpolation on the metrics is obtained. The proposed method is faster than those mentioned above and provides a continuous metric field with low variations. Some numerical examples illustrate our methodology.

3.5. Large deformation simulation using adaptive remeshing

Participants: Patrick Laug [correspondant], Houman Borouchaki.

The object of non-linear solid and structural mechanics is the modeling and the computation of structures with strong non-linearities, both geometrical and physical. The aim is to simulate the behavior of a mechanical part submitted to various mechanical stresses, in order to improve its mechanical strength, or even to optimize its manufacturing process with respect to damage occurrence. Among various theoretical, numerical and geometric tools involved in such a simulation, the interest in adaptive remeshing is really high nowadays. It is generally based on local refinement (governed by error estimation) and vertex smoothing strategies. Let us mention that the main difficulty lies in the fact that, in large strains, the domain geometry is variable and cannot be defined in an explicit way.

New contributions to the strategy using adaptive meshing and *a posteriori* error estimation in large elasto-plasticity have been developed. We are interested in the problem of remeshing a mechanical structure composed of several parts (which are in contact) subjected to large plastic deformations. A general scheme, constituted by several steps necessary to an almost optimal representation of the evolving domain, is proposed. These steps are divided into two main categories: the definition of the boundary of the deformed parts and the whole remeshing of the parts. The remeshing is governed by a mesh size map representing the conformity with the underlying geometry of the deformed parts, the improvement of the accuracy of the desired mechanical fields, and the convergence of the mechanical process as well. This size map results from an *a posteriori* estimation of the "interpolation error" independently from the considered mechanical fields. The final deformation after the whole simulation is assumed to be obtained iteratively by "small" deformations (which is the case in the framework of an explicit integration scheme to solve the problem). After such a small deformation, rigid parts are moved and deformable parts are slightly distorted (assuming that each mesh element is still valid). The remeshing is applied to deformable parts after each deformation increment. The proposed technique is used to simulate the impact of a projectile on a confined explosive. We show in particular that the ignition of the explosive appears in two different areas.

3.6. Maillage d'un milieu géologique et d'ouvrages de stockage

Participants: Patrick Laug [correspondant], Houman Borouchaki.

Cette étude a été menée dans le cadre du partenariat stratégique ANDRA/Inria. L'objectif est la construction d'un maillage statique 3D prenant en compte la géométrie des couches d'un milieu géologique et celle d'ouvrages de stockage afin de réaliser un calcul d'hydraulique et de transfert de solutés. En particulier, ce maillage sera exploité pour mener des calculs préparatoires aux calculs de sûreté. Il permettra de mieux représenter à l'échelle du milieu géologique les différentes voies de transfert (ouvrages et géologie multi-couches) des radionucléides, en considérant les évolutions géodynamiques, et de contribuer à identifier les simplifications éventuelles qui seront définies pour établir le modèle conceptuel de calcul de performances et de sûreté.

Les données d'entrée représentent la description géométrique du milieu géologique incluant les ouvrages de stockage. Le schéma de construction comprend quatre étapes :

1. Prétraitement des données d'entrée. Les sommets multiples du maillage volumique sont fusionnés afin de pouvoir extraire une topologie conforme. Grâce à cette topologie, les surfaces interfaces entre deux couches consécutives sont identifiées. Ces surfaces représentent des contraintes surfaciques que le mailleur volumique doit respecter. En outre, les lignes intersections entre ces surfaces contraintes, appelées lignes d'affleurement, sont identifiées. De même, ces lignes représentent des contraintes linéiques pour le mailleur volumique. Afin de définir la ligne polygonale associée à chaque rivière, les arêtes de l'enveloppe supérieure du maillage volumique de référence (surface topographique) dont les deux extrémités ont le même code de rivière sont identifiées.

2. Définition de la géométrie du domaine 2D de référence. On définit le plan de référence comme étant le plan d'équation $z = 0$, et le domaine 2D de référence comme la trace du polygone de l'extension horizontale dans ce plan. Toutes les contraintes linéiques (lignes d'affleurement, rivières et contours des ouvrages) sont projetées verticalement sur le plan de référence et leurs traces dans le domaine de référence sont retenues. En outre, des nouvelles lignes contraintes parallèles aux contours des ouvrages sont insérées afin de mieux contrôler la génération du maillage du domaine de référence. L'ensemble de toutes les lignes du domaine de référence est rendu conforme par ajout des points aux intersections éventuelles de ces lignes, et aussi par fusion des points et des lignes coïncidents.

3. Construction du maillage quad-dominant du domaine 2D de référence. Le maillage du domaine de référence est généré en utilisant un schéma adaptatif de construction de maillages quad-dominants. Dans un premier temps, un maillage quad-dominant initial du domaine est construit en spécifiant une taille fixe sur les lignes d'affleurement et les rivières et une taille dépendant de la grandeur des ouvrages sur ces derniers. Afin de contrôler la gradation du maillage (rapport maximal entre les longueurs d'arêtes issues d'un même sommet), deux maillages quad-dominants adaptés sont générés. Ici, l'adaptation consiste à modifier la carte de taille courante pour respecter le seuil de gradation spécifié.

4. Construction du maillage hex-dominant 3D du milieu. Le maillage volumique du milieu géologique est généré par extrusion verticale du maillage quad-dominant du domaine de référence. Deux types de configuration sont considérés : extrusion d'un quadrilatère (dit de base) du maillage du domaine de référence et extrusion d'un triangle (dit de base) du maillage du domaine de référence. Dans le premier cas, selon la configuration des surfaces (surfaces interfaces entre deux couches ou faces supérieures ou inférieures d'ouvrages) rencontrées, des hexaèdres et des prismes sont générés. Plus précisément, dans ce cas, l'extrusion résulte en un ensemble de quadrilatères ordonnés verticalement avec quatre arêtes appartenant à la même surface ou deux arêtes opposées appartenant chacune à une surface. Les quadrilatères consécutifs sont connectés et, en fonction du nombre de sommets communs entre deux quadrilatères consécutifs, des hexaèdres ou des prismes sont générés. Par ailleurs, une configuration de quadrilatère est validée si d'une part chaque élément résultant est géométriquement valide (hexaèdre, prisme ou pyramide) et si, d'autre part, il contient son barycentre et ses faces sont quasi-planes. Dans le cas contraire, le quadrilatère de base est subdivisé en deux triangles et généralement selon la diagonale donnant une configuration de deux triangles de Delaunay.

3.7. Advanced meshing and remeshing procedure for mechanical and numerical simulations

Participants: Abel Cherouat [correspondant], Houman Borouchaki, Paul-Louis George, Patrick Laug, Zhu Aichun, Jie Zhang, Faouzi Slimani, Guillaume Dufaye.

Most metal forming parts involve complex geometry and flow characteristics as large (visco)-plasticity flow, heat exchange, ductile damage, evolving contact with friction. An intrinsic difficulty in metal forming process is the constantly changing configuration of the deforming part (finite transformation, thermo-plastic flow). In metal forming, the mesh size should be adapted to the curvature of complex tools in order to optimize the contact boundaries and the damaged zones. These problems can be resolved if an adaptive remeshing scheme is incorporated automatically in the finite element analysis. It is necessary to adapt the mesh in order to improve the geometry of the deformed part and the damage localization. To mesh the 3D computational domain, we apply a new optimization approach which uses a combined Delaunay-frontal method to define field points and to construct the connection between these points or with a given prescribed size map (error estimate). The first objective of this project is to develop a 3D advanced remeshing procedure (error estimation, field transfer, optimisation meshing) for metal forming. The second objective is to integrate in a computational environment the mechanical model, 3D reconstruction from images, reliability-optimisation and the remeshing procedure using the ABAQUS/Explicit solver and the adaptive mesher. Application is dedicated to some examples (side pressing, blanking and orthogonal cutting, 3D guillotining, thermo-hydroforming and forging) for metal forming and breast and porous metal foam material reconstitution.

3.8. Effect of fibre geometrical morphology on the mechanical properties of PolyPropylene Hemp fibre composite material

Participants: Abel Cherouat [correspondant], Florent Ilczyszyn.

These last years, hemp fibres have been used as reinforcement for compound based on polymer in different industrial manufacturing for their interesting mechanical and ecological properties. Hemp fibres present a non-homogeneous cross section and complex geometry that can have a high effect on their mechanical properties. The mechanical properties of hemp fibres are rather difficult to determine and request a specific characterization method. In this project, micro-tensile tests coupled with numerical imaging treatments, meshing reconstitution and finite elements computations are investigated. The numerical imaging allows to define finely the hemp cross section along the fibre and aims to reconstruct a 3D hemp fibre CAD using adaptive mesh.

3.9. Mise au point de méthodes de remaillage adaptatif 3D dans le cadre de simulations numériques de mise en forme de structure minces

Participants: Houman Borouchaki, Abel Cherouat, Laurence Moreau [correspondant].

Au cours des simulations numériques de mise en forme en 3D, les grandes déformations mises en jeu font que le maillage subit de fortes distorsions. Il est alors nécessaire de remailler continuellement la pièce afin de pouvoir capturer les détails géométriques des surface en contact, adapter la taille du maillage à la solution physique et surtout pouvoir effectuer la simulation jusqu'à la fin du procédé de mise en forme. Lorsque la pièce est comprise entre des outils rigides (cas de l'emboutissage), aux problèmes de remaillage s'ajoutent aussi des difficultés sur la gestion du contact entre les pièce. Une méthode couplant une stratégie de remaillage adaptatif et une technique de projection a été développée. La méthode de remaillage adaptatif, basée sur des techniques de raffinement et déraffinement est contrôlée par des cartes de taille géométrique et physique. La projection des nouveaux nœuds sur l'outil permet de conserver le contact entre la pièce et l'outil. Afin de pouvoir réaliser des simulations numériques de composites tissés, une procédure spécifique a été ajoutée au remailleur afin de pouvoir raffiner les éléments finis bi-composants (association d'éléments finis de barre et de membrane orientés matérialisant le comportement de fibres chaîne et trame). Le formage incrémental est un procédé de mise en forme de tôle récent sans poinçon ni matrice, basé sur la déformation progressive du flan à l'aide d'un simple outil de forme hémisphérique commandé par une machine à commande numérique. L'inconvénient de ce nouveau procédé étant le temps de calcul, nous avons proposé une méthode de remaillage adaptatif permettant de raffiner le maillage uniquement au voisinage de l'outil rigide, là où les déformations ont lieux et permettant de déraffiner le maillage après le passage de l'outil rigide.

3.10. Mise au point de méthodes de remaillage adaptatif 3D dans le cadre de simulations numériques de mise en forme de structure minces

Participants: Houman Borouchaki, Abel Cherouat, Laurence Moreau [correspondant].

L'objectif est de reconstruire un maillage de la surface 3D d'un buste féminin à partir d'images 2D issues des prises de vue simultanées de plusieurs appareils photos numériques (photos prises sous des angles différents). Une cabine de mesure, équipée de 24 appareils photos numériques, 6 vidéoprojecteurs, pilotée par un ordinateur extérieur à la cabine a été développée et permet d'acquérir de manière simultanée 24 photos numériques du buste sous des angles différents. Un algorithme original basé sur l'utilisation d'un motif projeté sur le buste a été développé et programmé pour la corrélation entre les images 2D. Une méthode de triangulation 3D associée à une technique d'optimisation a été développée et permet de déterminer les positions 3D des points à partir des pixels de vues différentes.

3.11. Applications du maillage et développements de méthodes avancées pour la cryptographie

Participants: Dominique Barchiesi [correspondant], Thomas Grosges, Michael François.

L'utilisation des nombres (pseudo)-aléatoires a pris une dimension importante ces dernières décennies. De nombreuses applications dans le domaine des télécommunications, de la cryptographie, des simulations numériques ou encore des jeux de hasard, ont contribué au développement et à l'usage de ces nombres. Les méthodes utilisées pour la génération de tels nombres (pseudo)-aléatoires proviennent de deux types de processus : physique et algorithmique. Ce projet de recherche a donc pour objectif principal le développement de nouveaux procédés de génération de clés de chiffrement, dits "exotiques", basés sur des processus physiques, multi-échelles, multi-domaines assurant un niveau élevé de sécurité. Deux classes de générateurs basés sur des principes de mesures physiques et des processus mathématiques ont été développé.

La première classe de générateurs exploite la réponse d'un système physique servant de source pour la génération des séquences aléatoires. Cette classe utilise aussi bien des résultats de simulation que des résultats de mesures interférométriques pour produire des séquences de nombres aléatoires. L'application du maillage adaptatif sert au contrôle de l'erreur sur la solution des champs physiques (simulés ou mesurés). A partir de ces cartes physiques, un maillage avec estimateur d'erreur sur l'entropie du système est appliqué. Celui-ci permet de redistribuer les positions spatiales des noeuds. L'étude (locale) de la réduction d'entropie des clés tout au long de la chaîne de création et l'étude (globale) de l'entropie de l'espace des clés générées sont réalisées à partir de tests statistiques.

La seconde classe de générateurs porte sur le développement de méthodes avancées et est basée sur l'exploitation de fonctions chaotiques en utilisant les sorties de ces fonctions comme indice de permutation sur un vecteur initial. Ce projet s'intéresse également aux systèmes de chiffrement pour la protection des données et deux algorithmes de chiffrement d'images utilisant des fonctions chaotiques sont développés et analysés. Ces Algorithmes utilisent un processus de permutation-substitution sur les bits de l'image originale. Une analyse statistique approfondie confirme la pertinence des cryptosystèmes développés.

3.12. Développement de méthodes avancées et maillages appliqués àa l'étude de la nanomorphologie des nanotubes/fils en suspension liquide

Participants: Dominique Barchiesi, Houman Borouchaki, Abel Cherouat, Anis Chaari, Thomas Grosges [correspondant], Laurence Moreau.

Ce projet de recherche (NANOMORPH) a pour objet principal le de´veloppement et la mise au point d'une instrumentation optique pour de´terminer la distribution en tailles et le coefficient de forme de nanofils (NF) ou de nanotubes (NT) en suspension dans un e´coulement. Au cours de ce projet, deux types de techniques optiques comple´mentaires sont de´veloppe´es. La premie`re, base´e sur la diffusion statique de la lumie`re, ne´cessite d'e´tudier au pre´alable la physico-chimie de la dispersion, la stabilisation et l'orientation des nanofils dans les milieux d'e´tude. La seconde me´thode, base´e sur une me´thode opto-photothermique pulse´e, ne´cessite en sus, la mode´lisation de l'interaction laser/nanofils, ainsi que l'e´tude des phe´nome`nes multiphysiques induits par ce pro-cessus. L'implication de l'e´quipe-projet GAMMA3 concerne principalement la simulation mul- tiphysique de l'interaction laser-nanofils et l'e´volution temporelle des bulles et leurs formations. L'une des principales difficulte´s de ces proble´matiques est que la ge´ome´trie du domaine est variable (a` la fois au sens ge´ome´trique et topologique). Ces simulations ne peuvent donc eˆtre re´alise´es que dans un sche´ma adaptatif de calcul ne´cessitant le remaillage tridimensionnel mobile, de´formable avec topologie variable du domaine (formation et e´volution des bulles au cours du temps et de l'espace).

3.13. Applications du maillage à des problèmes multi-physiques, développement de méthodes de résolutions avancées et modélisation électromagnetisme-thermique-mécanique à l'échelle mesoscopique

Participants: Dominique Barchiesi [correspondant], Thomas Grosges, Abel Cherouat, Thomas Grosges, Houman Borouchaki, Laurence Giraud-Moreau, Sameh Kessentini, Anis Chaari, Fadhil Mezghani.

Le contrôle et l'adaptation du maillage lors de la résolution de problèmes couplés ou/et non linéaires reste un problème ouvert et fortement dépendant du type de couplage physique entre les EDP à résoudre. Notre objectif est de développer des modèles stables afin de calculer les dilatations induites par l'absorption d'énergie électromagnétique, par des structures matérielles inférieures au micron. Les structures étudiées sont en particulier des nanoparticules métalliques en condition de résonance plasmon. Dans ce cas, un maximum d'énergie absorbée est attendu, accompagné d'un maximum d'élévation de température et de dilatation. Il faut en particulier développer des modèles permettant de simuler le comportement multiphysique de particules de formes quelconques, pour une gamme de fréquences du laser d'éclairage assez étendue afin d'obtenir une étude spectroscopique de la température et de la dilatation.L'objectif intermédiaire est de pouvoir quantifier la dilatation en fonction de la puissance laser incidente. Le calcul doit donc être dimensionné et permettre finalement des applications dans les domaines des capteurs et de l'ingénierie biomédicale. En effet, ces nanoparticules métalliques sont utilisées à la fois pour le traitement des cancers superficiels par nécrose de tumeur sous éclairage adéquat, dans la fenêtres de transparence cellulaire. Déposées sur un substrat de verre, ces nanoparticules permettent de construire des capteurs utilisant la résonance plasmon pour être plus sensibles (voir projet européen *Nanoantenna* et l'activité génération de nombres aléatoires. Cependant, dans les deux cas, il est nécessaire, en environnement complexe de déterminer la température locale, voire la dilatation de ces nanoparticules, pouvant conduire à une désaccord du capteur, la résonance plasmon étant très sensible aux paramètres géométriques et matériels des nanostructures. Dans ce sens, l'étude permet d'aller plus loin que la << simple >> interaction électromagnétique avec la matière du projet européen *Nanoantenna*.

Le travail de l'année 2012 a constitué en une pré-étude des spécificités de ce type de problème multiphysique pour des structures de forme simple et la mise en place de fonctions test de référence, pour les développements de maillage adaptatifs pour les modèles multiphysiques éléments finis. Nous espérons pouvoir proposer un projet ANR couplant les points de vue microscopiques et macroscopiques dans les deux années qui viennent.

3.14. Mesh adaptation for very high-order numerical scheme

Participants: Frederic Alauzet [correspondant], Adrien Loseille, Estelle Mbinky.

In the past, we have demonstrate that multi-scale anisotropic mesh adaptation is a powerful tool to accurately simulate compressible flow problem and to obtain faster convergence to continuous solutions. But, this was limited to second order numerical scheme. Nowadays, numerous teams are working on the development of very high-order numerical scheme (e.g. of third or greater order): Discontinous Galerkin, Residual Distribution scheme, Spectral method, ...

This work extend interpolation error estimates to higher order numerical solution representation. We have examined the case of third-order accuracy. The first step is to reduce the tri-linear form given by the third order error term into a quadratic form based on the third order derivative. From this local error model, the optimal mesh is exhibited thanks to the continuous mesh framework.

3.15. Visualisation et modification des maillages courbes d'ordre élevé

Participants: Julien Castelneau, Adrien Loseille [correspondant], Loïc Maréchal.

Dans le cadre du projet ILab, des nouveaux algorithmes de visualisation et de modifications interactives des maillages courbes et hybrides ont été développés. En effet, une des principales difficultés dans la génération de maillages courbes reste la visualisation. Il est également nécessaire de disposer d'algorithmes de corrections interactifs car les maillages de surfaces initiaux (de degré 2) sont pour la plupart faux.

3.16. A changing-topology ALE numerical scheme

Participants: Frédéric Alauzet [correspondant], Nicolas Baral.

The main difficulty arising in numerical simulations with moving geometries is to handle the displacement of the domain boundaries, *i.e.,* the moving bodies. Only vertices displacement is not sufficient to achieve complex movement such as shear. We proved that the use of edge swapping allows us to achieve such complex displacement. We therefore developed an ALE formulation of this topological mesh modification to preserve the solver accuracy and convergence order. The goal is to extend to 3D the previous work done in 2D.

3.17. Mesh adaptation for Navier-Stokes Equations

Participants: Frédéric Alauzet, Victorien Menier, Adrien Loseille [correspondant].

Adaptive simulations for Navier-Stokes equations require to propose accurate error estimates and design robust mesh adaptation algorithms (for boundary layers).

For error estimates, we design new estimates suited to accurately capture the speed profile in the boundary layers. For mesh adaptation, we design a new method to generate structured boundary layer meshes which are mandatory to accurately compute compressible flows a high Reynolds number (several millions). It couple the specification of the optimal boundary layer from the geometry boundary and moving mesh techniques to extrude the boundary layer in an already existing mesh. The main advantage of this approach is its robustness, *i.e.,* at each step of the algorithm we have always a valid mesh.

3.18. Maillages hexaédriques et calcul parallèle

Participant: Loïc Maréchal [correspondant].

Développement d'un remailleur de surfaces par la méthode octree. Celui-ci permet de passer d'une surface triangulée à problèmes (intersections de triangles, non-conformités, trous, etc.) à un maillage valide au sens des éléments finis.

Nouvelle version de la librairie d'aide au calcul sur GPU, GMLIB2, permettant de porter des codes travaillant sur des maillages de manière bien plus simple et efficace que la précédente. Des accélérations de l'ordre de 30, par rapport à un CPU en séquentiel, ont été obtenus avec le solveur Wolf et le mailleur Hexotic sur une carte Quadro 6000.

De nombreux développements sur le mailleur hexaédrique Hexotic ont été réalisés suite aux demandes de nombreux acheteurs industriels potentiels.

GECO Team

6. New Results

6.1. New results: geometric control

We start by presenting some results on the design of motion planning and tracking algorithms.

- In [10] we present an iterative steering algorithm for nonholonomic systems (also called driftless control-affine systems) and we prove its global convergence under the sole assumption that the Lie Algebraic Rank Condition (LARC) holds true everywhere. That algorithm is an extension of the one introduced in [65] for regular systems. The first novelty here consists in the explicit algebraic construction, starting from the original control system, of a lifted control system which is regular. The second contribution of the paper is an exact motion planning method for nilpotent systems, which makes use of sinusoidal control laws and which is a generalization of the algorithm described in [83] for chained-form systems.
- [6] and [5] are about motion planning for kinematic systems, and more particularly ε-approximations of non-admissible trajectories by admissible ones. This is done in a certain optimal sense. The resolution of this motion planing problem is showcased through the thorough treatment of the ball with a trailer kinematic system, which is a non-holonomic system with flag of type $(2, 3, 5, 6)$.

Application-oriented results about motion planning are contained in [15]. The paper proposes in particular a strategy for providing Unmanned Aerial Vehicles with a certain degree of autonomy, via autonomous planification/replanification strategies.

Let us list some new results in sub-Riemannian geometry and hypoellitpic diffusion.

- In [1] we study the Radon-Nikodym derivative of the spherical Hausdorff measure with respect to a smooth volume for a regular sub-Riemannian manifold . We prove that this is the volume of the unit ball in the nilpotent approximation and it is always a continuous function. We then prove that up to dimension 4 it is smooth, while starting from dimension 5, in corank 1 case, it is C^3 (and C^4 on every smooth curve) but in general not C^5. These results answer to a question addressed by Montgomery about the relation between two intrinsic volumes that can be defined in a sub-Riemannian manifold, namely the Popp and the Hausdorff volume. If the nilpotent approximation depends on the point (that may happen starting from dimension 5), then they are not proportional, in general.
- In [9] we study the Laplace–Beltrami operator on generalized Riemannian structures on orientable surfaces for which a local orthonormal frame is given by a pair of vector fields that can become collinear. Under the assumption that the structure is 2-step Lie bracket generating, we prove that the Laplace–Beltrami operator is essentially self-adjoint and has discrete spectrum. As a consequence, a quantum particle cannot cross the singular set (i.e., the set where the vector fields become collinear) and the heat cannot flow through the singularity.
- For an equiregular sub-Riemannian manifold M, Popp's volume is a smooth volume which is canonically associated with the sub-Riemannian structure, and it is a natural generalization of the Riemannian one. In [4] we prove a general formula for Popp's volume, written in terms of a frame adapted to the sub-Riemannian distribution. As a first application of this result, we prove an explicit formula for the canonical sub-Laplacian, namely the one associated with Popp's volume. Finally, we discuss sub-Riemannian isometries, and we prove that they preserve Popp's volume. We also show that, under some hypotheses on the action of the isometry group of M, Popp's volume is essentially the unique volume with such a property.

- In [21], for a sub-Riemannian manifold provided with a smooth volume, we relate the small time asymptotics of the heat kernel at a point y of the cut locus from x with roughly "how much" y is conjugate to x. This is done under the hypothesis that all minimizers connecting x to y are strongly normal, i.e. all pieces of the trajectory are not abnormal. Our result is a refinement of the one of Leandre $4t \log p_t(x, y) \to -d^2(x, y)$ for $t \to 0$, in which only the leading exponential term is detected. Our results are obtained by extending an idea of Molchanov from the Riemannian to the sub-Riemannian case, and some details we get appear to be new even in the Riemannian context. These results permit us to obtain properties of the sub-Riemannian distance starting from those of the heat kernel and vice versa. For the Grushin plane endowed with the Euclidean volume we get the expansion $p_t(x, y) \sim t^{-5/4} \exp(-d^2(x, y)/4t)$ where y is reached from a Riemannian point x by a minimizing geodesic which is conjugate at y. In [22] we investigate the small time heat kernel asymptotics on the cut locus on the class of two-spheres of revolution, which is the simplest class of 2-dimensional Riemannian manifolds different from the sphere with nontrivial cut-conjugate locus. We determine the degeneracy of the exponential map near a cut-conjugate point and present the consequences of this result to the small time heat kernel asymptotics at this point. These results give a first example where the minimal degeneration of the asymptotic expansion at the cut locus is attained.
- In [24] we studied normal forms for 2-dimensional almost-Riemannian structures. The latter are generalized Riemannian structures on surfaces for which a local orthonormal frame is given by a Lie bracket generating pair of vector fields that can become collinear. Generically, there are three types of points: Riemannian points where the two vector fields are linearly independent, Grushin points where the two vector fields are collinear but their Lie bracket is not, and tangency points where the two vector fields and their Lie bracket are collinear and the missing direction is obtained with one more bracket. In [24] we consider the problem of finding normal forms and functional invariants at each type of point. We also require that functional invariants are complete, in the sense that they permit to recognize locally isometric structures. The problem happens to be equivalent to the one of finding a smooth canonical parameterized curve passing through the point and being transversal to the distribution. For Riemannian points such that the gradient of the Gaussian curvature K is different from zero, we use the level set of K as support of the parameterized curve. For Riemannian points such that the gradient of the curvature vanishes (and under additional generic conditions), we use a curve which is found by looking for crests and valleys of the curvature. For Grushin points we use the set where the vector fields are parallel. Tangency points are the most complicated to deal with. The cut locus from the tangency point is not a good candidate as canonical parameterized curve since it is known to be non-smooth. Thus, we analyse the cut locus from the singular set and we prove that it is not smooth either. A good candidate happens to be a curve which is found by looking for crests and valleys of the Gaussian curvature. We prove that the support of such a curve is uniquely determined and has a canonical parametrization.

6.2. New results: quantum control

New results have been obtained for the control of the bilinear Schrödinger equation.

- In [16] we obtained a sufficient condition for approximate controllability of the bilinear discrete-spectrum Schrödinger equation exploiting the use of more than one control. The controllability result extends to simultaneous controllability, approximate controllability in H^s, and tracking in modulus. The result is more general than those present in the literature even in the case of one control and permits to treat situations in which the spectrum of the uncontrolled operator is very degenerate (e.g. multiple eigenvalues or presence of equal gaps among eigenvalues). These results are applied to the case of a rotating polar linear molecule in the space, driven by three external fields. A remarkable property of this model is the presence of infinitely many degeneracies and resonances in the spectrum preventing the application of the results in the literature.
- In [19] we present a constructive method to control the bilinear Schrödinger equation by means of

two or three controlled external fields. The method is based on adiabatic techniques and works if the spectrum of the Hamiltonian admits eigenvalue intersections, with respect to variations of the controls, and if the latter are conical. We provide sharp estimates of the relation between the error and the controllability time.

- In [18] we consider the minimum time population transfer problem for a two level quantum system driven by two external fields with bounded amplitude. The controls are modeled as real functions and we do not use the Rotating Wave Approximation. After projection on the Bloch sphere, we tackle the time-optimal control problem with techniques of optimal synthesis on 2-D manifolds. Based on the Pontryagin Maximum Principle, we characterize a restricted set of candidate optimal trajectories. Properties on this set, crucial for complete optimal synthesis, are illustrated by numerical simulations. Furthermore, when the two controls have the same bound and this bound is small with respect to the difference of the two energy levels, we get a complete optimal synthesis up to a small neighborhood of the antipodal point of the starting point.

6.3. New results: neurophysiology

- In [17] we study the global properties of an optimal control model of geometry of vision due to Petitot, Citti and Sarti. In particular, we consider the problem of minimizing $\int_0^L \sqrt{\xi^2 + K^2(s)}\, ds$ for a planar curve having fixed initial and final positions and directions. The total length L is free. Here s is the variable of arclength parametrization, $K(s)$ is the curvature of the curve and $\xi > 0$ a parameter. The main feature of the problem is that, if for a certain choice of boundary conditions there exists a minimizer, then this minimizer is smooth and has no cusp. However, not for all choices of boundary conditions there is a global minimizer. We study existence of local and global minimizers for this problem. We prove that if for a certain choice of boundary conditions there is no global minimizer, then there is neither a local minimizer nor a stationary curve (geodesic). We give properties of the set of boundary conditions for which there exists a solution to the problem. Finally, we present numerical computations of this set.
- In [2] we studied the general problem of reconstructing the cost from the observation of trajectories, in a problem of optimal control. It is motivated by the problem of determining what is the cost minimized in human locomotion. This applied question is very similar to the following applied problem, concerning HALE drones: one would like them to decide by themselves for their trajectories, and to behave at least as a good human pilot. These starting points are the reasons for the particular classes of control systems and of costs under consideration. To summarize, our conclusion is that in general, inside these classes, three experiments visiting the same values of the control are needed to reconstruct the cost, and two experiments are in general not enough. The method is constructive. The proof of these results is mostly based upon the Thom's transversality theory.

6.4. New results: switched systems

- In [12] we study the phenomenon of polynomial instability of switched systems. Stability properties for continuous-time linear switched systems are at first determined by the (largest) Lyapunov exponent associated with the system, which is the analogue of the joint spectral radius for the discrete-time case. We provided a characterization of marginally unstable systems, i.e., systems for which the Lyapunov exponent is equal to zero and such that there exists an unbounded trajectory. We also analyzed the asymptotic behavior of their trajectories. Our main contribution consists in pointing out a resonance phenomenon associated with marginal instability. In the course of our study, we derived an upper bound of the state at time t, which is polynomial in t and whose degree is computed from the resonance structure of the system. We also derived analogous results for discrete-time linear switched systems.
- The paper [13] is concerned with the stability of planar linear singularly perturbed switched systems of the type $\dot{x}(t) = \sigma(t)A_1^{\epsilon}x(t) + (1-\sigma(t))A_2^{\epsilon}x(t)$, where $\sigma : [0, +\infty) \to \{0, 1\}$, A_1^{ϵ} and A_2^{ϵ} are real matrices which represent singularly perturbed modes. By ϵ we denote here the parameter of

singular perturbation. We propose a characterization of the stability properties of such singularly perturbed switched systems based on the results given in [47]. More generally, we study transitions as ϵ varies and we restrict their number and nature. Finally, we compare the results obtained in this way with the Tikhonov-type results for differential inclusions obtained in the literature.

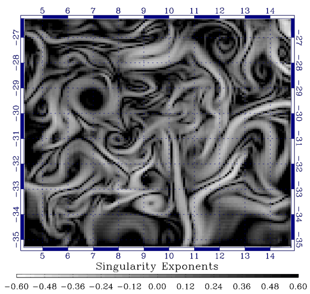
GEOSTAT Project-Team

6. New Results

6.1. Multiresolution analysis and optimal inference for high resolution ocean dynamics and ocean/atmosphere fluxes

Participants: Hussein Yahia [correspondant], Véronique Garçon, Oriol Pont, Joel Sudre, Christine Provost, Antonio Turiel, Christoph Garbe, Claire Pottier, Boris Dewitte.

A $p_{CO_2}^{ocean}$ signal computed as an output from the ROMS coupled physical/biogeochemical simulation model possesses the characteristics of the presence of a multiscale organization, typical of turbulence, which can be evidenced by the computation of singularity spectra. The multiscale organization is related to the cascading properties of intensive variables acquired from the underlying system. We show how to perform inference along the scales in order to build higher resolution of $p_{CO_2}^{ocean}$ maps. Figure 4 illustrates clearly one of the main ideas implemented in this study: coherent structures of $p_{CO_2}^{ocean}$ and SST (Sea Surface Temperature) signals are related, and the LPEs, which are dimensionless quantities recording transition strengths in a signal, encode properly the multiscale transitions.



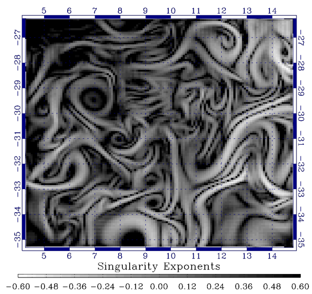


Figure 4. Local Predictability Exponents (LPEs) of ROMS-simulated $p_{CO_2}^{ocean}$ signal (left) and of corresponding SST (Sea Surface Temperature) generated signal (right). Transitions are are visually and quantitatively correlated, although not the same.

We perform a linear regression test:

$$\mathcal{S}(p_{CO_2}^{ocean})(x) = a(x)\mathcal{S}(SST)(x) + b(x)\mathcal{S}(CHLa)(x) + c(x) \tag{4}$$

with $\mathcal{S}(p_{CO_2}^{ocean})(x)$: LPE of $p_{CO_2}^{ocean}$ at x, $\mathcal{S}(SST)(x)$: LPE of SST at x, $\mathcal{S}(CHLa)(x)$: LPE of CHLa signal at x (CHLa: ocean colour data, corresponding to chlorophyl concentration). Tests are conducted over a period of 10 years on ROMS simulated data, with images corresponding to 128×128 pixels for the high resolution and 32×32 for the low resolution. There is one data every 10 days. In figure 5 we compare the functional dependencies of $p_{CO_2}^{ocean}$ vs. SST and CHLa with those of the corresponding LPEs: the original signals are physical variables of different dimensions, with complex undetermined functional dependencies. On the contrary, the dimensionless LPEs of these variables, which record the multiscale transitions, display clearly a much simpler dependency, approximated at satisfactory precision by a linear regression.

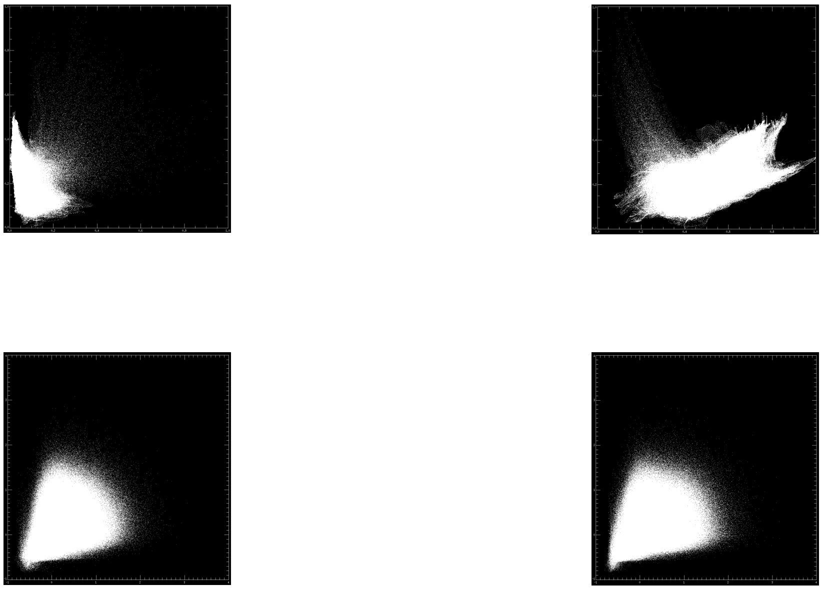


Figure 5. Pictures indicating the nature of the functional dependencies of $p_{CO_2}^{ocean}$ vs. CHLa (top left), of $p_{CO_2}^{ocean}$ vs. SST (top right), of $\mathcal{S}(p_{CO_2}^{ocean})$ vs $\mathcal{S}(CHLa)$ (bottom left) and of $\mathcal{S}(p_{CO_2}^{ocean})$ vs $\mathcal{S}(SST)$ (bottom right). The dependencies are computed on a 10-year ROMS simulation dataset, with a time frequency of one every 10 days.

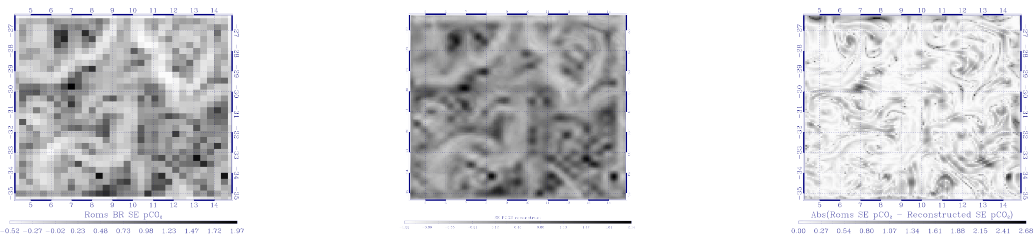


Figure 6. Left : the low resolution version of LPEs for $\mathcal{S}(p_{CO_2}^{ocean})$. Middle: result of the reconstruction. Right: absolute difference map between the ROMS generated high resolution LPEs and the reconstructed.

We prove the feasibility of a reconstruction by computing the high resolution LPEs $\mathcal{S}(p_{CO_2}^{ocean})(x)$ from their low resolution counterparts and an effective multiresolution analysis, using only an approximation of the optimal wavelet in the form of a Battle-Lemarié 3-31 mother wavelet. We show in figure 6 the results obtained by inference along the scales. The reconstructed LPEs of $p_{CO_2}^{ocean}$ are in good correspondence with the original high resolution signal.

- Related publications: [15], [16], [24], [14].

6.2. Singularity analysis and reconstructible systems

Participants: Oriol Pont [correspondant], Hussein Yahia, Antonio Turiel.

The local singularity exponents of a signal are directly related to the distribution of information in it. This fact implies that accurate evaluation of such exponents opens the door to signal reconstruction and characterisation of the dynamical parameters of the process originating the signal. Many practical implications arise in a context of digital signal processing, since the information on singularity exponents is usable for compact encoding, reconstruction and inference. The evaluation of singularity exponents in a digital context is not straightforward and requires the calculation of the Unpredictable Point Manifold of the signal. In this work, we present an algorithm for estimating the values of singularity exponents at every point of a digital signal of any dimension. We show that the key ingredient for robust and accurate reconstructibility performance lies on the definition of multiscale measures in the sense that they encode the degree of singularity and the local predictability at the same time. See figure 7 .

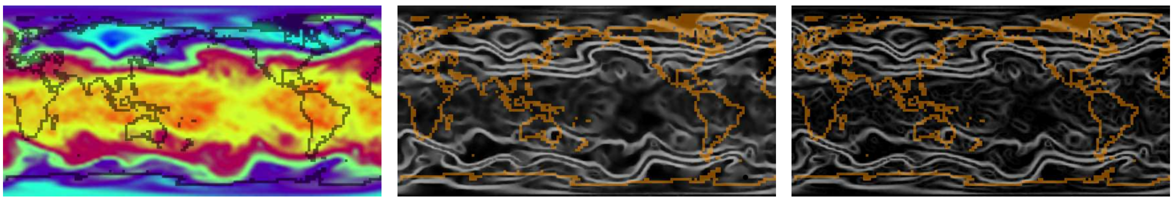


Figure 7. Left: 876576th hour slice of ERA-40 artificially rescaled 4x with bicubic interpolation for the purpose of clarity of illustration. Middle: singularity exponents calculated only in the space. Some rescaling artefacts visibly appear but without significant disturbance of the fine structure details. Right: singularity exponents calculated in the space-time domain. Notice the increased degree of detail when the temporal information is taken into account.

- Related publication: [13].

6.3. Multiscale analysis of the heart electric potential: describing atrial fibrillation

Participants: Oriol Pont [correspondant], Hussein Yahia, Rémi Dubois.

The cardiac electrical activity is a complex system, for which nonlinear signal-processing is required to characterize it properly. In this context, an analysis in terms of singularity exponents is shown to provide compact and meaningful descriptors of the structure and dynamics. In particular, singularity components reconstruct the epicardial electric potential maps of human atria, inverse-mapped from surface potentials; such approach describe sinus-rhythm dynamics as well as atrial flutter and atrial fibrillation. See figure 2 .

- Related publications: [12], [20], [23].

6.4. Edges, transitions and criticality

Participants: Suman Maji [correspondant], Hussein Yahia.

In this work, various notions of edges encountered in digital image processing are reviewed in terms of compact representation (or completion). We show that critical exponents defined in Statistical Physics lead to a much more coherent definition of edges, consistent across the scales in acquisitions of natural phenomena, such as high resolution natural images or turbulent acquisitions. Edges belong to the multiscale hierarchy of an underlying dynamics, they are understood from a statistical perspective well adapted to fit the case of natural images. Numerical computation methods for the evaluation of critical exponents in the non-ergodic case are recalled, which apply for the vast majority of natural images. We study the framework of reconstructible systems in a microcanonical formulation, show how it redefines edge completion, and how it can be used to evaluate and assess quantitatively the adequation of edges as candidates for compact representations. We study with particular attention the case of turbulent data, in which edges in the classical sense are particularly challenged. Tests are conducted and evaluated on a standard database for natural images. We test the newly introduced compact representation as an ideal candidate for evaluating turbulent cascading properties of complex images, and we show better reconstruction performance than the classical tested methods. See figure 8
.

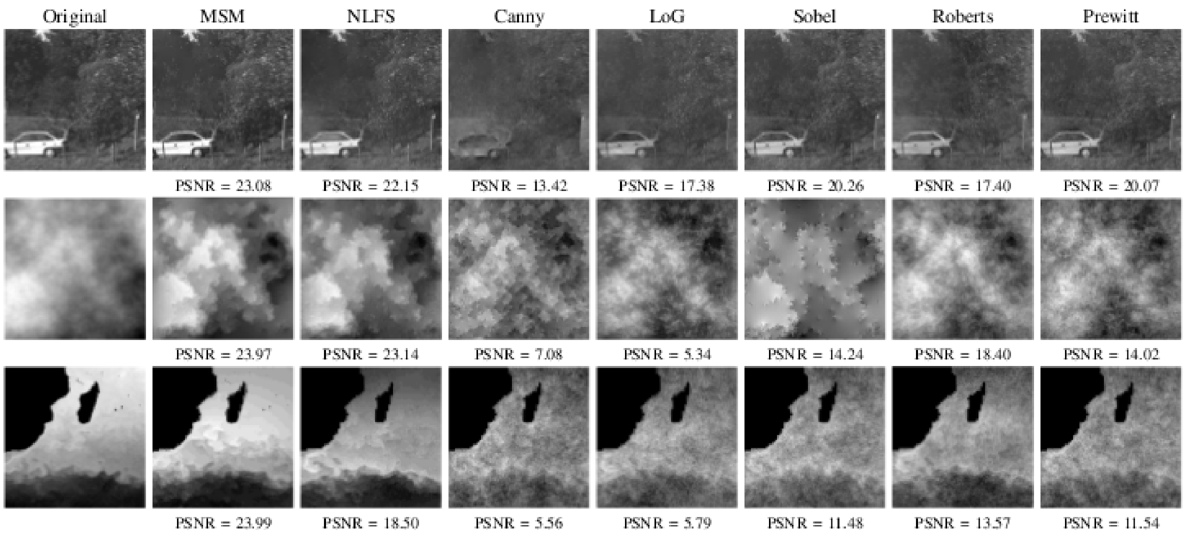


Figure 8. From left to right in each line: an original input image, and the reconstruction performed on the outputs resulting from various edge detection algorithms, showing the superiority of edge pixels computed from the Microcanonical Multiscale Formalism (column MSM). Note that NLFS [32], which is based on nonlinear filtering, performs the best after MSM.

6.5. Reconstruction of Optical phase from acquired sub-image gradients

Participants: Suman Maji [correspondant], Hussein Yahia, Thierry Fusco.

Turbulence in the Earth's atmosphere leads to a distortion in the planar wavefront from outer space resulting in a phase error. This phase error is responsible for the refractive blurring of images accounting to the loss in spatial resolution power of ground based telescopes. The common mechanism used to remove phase error from incoming wavefront is Adaptive Optics (AO). In AO systems, an estimate of the phase error is obtained from the gradient measurements of the wavefront collected by a Hartmann-Shack (HS) sensor. The correction estimate is then passed through a servo-control loop to a deformable mirror which compensates for the loss in resolution power. In this work, we propose a new approach to reconstructing the phase error from the HS

Table 1: Evaluation of the reconstructed phase under different levels of noise.

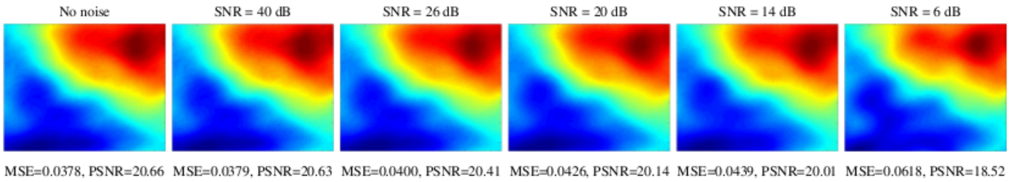


Table 2: Evaluation of the reconstructed phase using log power spectrum (**row 1**) and atmospheric structure functions (**row 2**).

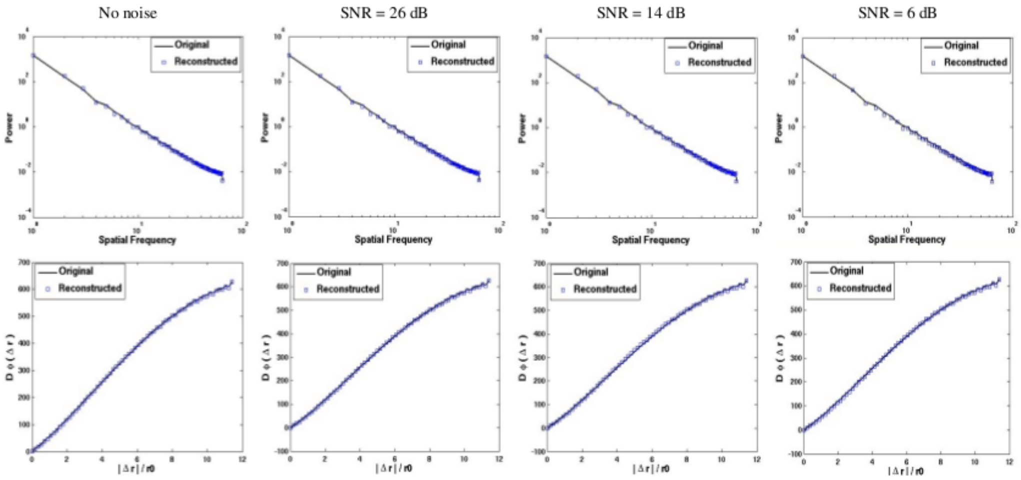


Figure 9. Results showing the robustness of the multiscale phase reconstruction algorithm for Adaptive Optics (AO) under various conditions of noise.

gradient measurements using the MMF. We also validate the results using standard validation techniques in Adaptive Optics (log power spectrum, structure functions). See figure 9 .

- Related publications: [18], [19].

6.6. Discriminative learning for Automatic speaker recognition

Participants: Reda Jourani [correspondant], Khalid Daoudi, Régine André-Obrecht, Driss Aboutajdine.

We continued our work aiming at developing efficient versions of Large Margin Gaussian Mixture Models (LM-GMM) for speaker identification. We developed a new and efficient learning algorithm and evaluated it on NIST-SRE'2006 data. The results show that, combined with the channel compesentation technique SFA, this new algorithm outperforms the state-of-the-art discriminative method GMM-supervectors SVM combined with NAP compensatation.

- Related publication: [10].

6.7. Speech Analysis

Participants: Vahid Khanagha [correspondant], Khalid Daoudi, Hussein Yahia, Oriol Pont.

- Development of a GCI detection algorithm (Vahid Khanagha, Khalid Daoudi, Hussein Yahia). According to the aerodynamic theory of voicing, the excitation source for voiced speech sounds is represented as glottal pulses, which to a first approximation, can be considered to occur at discrete instants of time. This major excitation usually coincides with the Glottal Closure Instants (the GCIs). The precise detection of GCIs has found many applications in speech technology: accurate estimation of vocal tract system, pitch marking of speech for pitch synchronous speech processing algorithms, conversion of pitch and duration of speech recordings, prosody modification and synthesis. We use the MMF for detection of these physically important instants. To do so, we study the correspondence of the Most Singular Manifold with the physical production mechanism of the speech signal and we show that this subset can be used for GCI detection. We show that, in clean speech, our algorithm has similar performance to recent methods and, in noisy speech, it significantly outperforms state-of-the-art methods. Indeed, as our algorithm is based on both time domain and inter-scale smoothings, it provides higher robustness against many types of noises. In the mean-time, the high geometrical resolution of singularity exponents prevents the accuracy to be compromised. Moreover, the algorithm extracts GCIs directly from the speech signal and does not rely on any model of the speech signal (such as the autoregressive model in linear predictive analysis). See figure 10 .
- Development of an efficient algorithm for sparse Linear Prediction Analysis (Vahid Khanagha, Khalid Daoudi). We address the problem of sparse Linear Prediction (LP) analysis, which involves the estimation of vocal tract model such that the corresponding LP residuals are as sparse as possible: for voiced sounds, one desires the residual to be zero all the time, except for few impulses at GCIs. Sparse Linear Prediction Analysis (LPA) problem has recently got much scientific attention and its classical solutions suffer from computational and algorithmic complexties. We introduce a simple closed-form solution in this chapter which is based on the minimization of weighted l_2-norm of residuals. The weighting function plays the most important role in our solution in maintaining the sparsity of the resulting residuals. We use our MSM-based GCI detector to extract from the speech signal itself, the points having the potential of attaining largest norms of residuals and then we construct the weighting function such that the prediction error is relaxed on these points. Consequently, the weighted l_2-norm objective function can be efficiently minimized by the solution of normal equations of liner least squares problem. The choice of our MSM-based GCI detector is particularly justified, considering the fact that most of the successful GCI detection methods actually use LP residuals for their detection and hence, they cannot be used for constraining the LP problem. Our algorithm is completely independent of any model that might be assumed for speech signal. We will see that when compared to classical techniques, our simple algorithm provides better sparseness

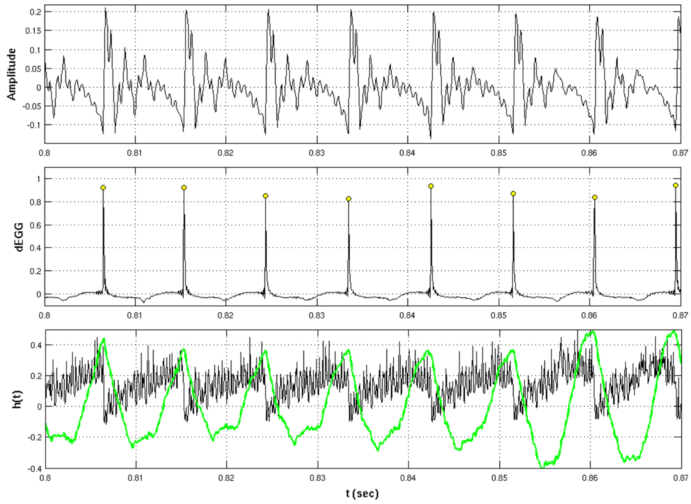


Figure 10. Top: a voiced segment of the speech signal taken from KED database. Middle: the differenced Electro-Glotto-Graph signal which serves for extraction of reference GCI points. The peaks are marked with yellow circles as the reference GCIs. Bottom: singularity exponents are shown by black color and an auxiliary functional showing changes in DC level of exponents is shown in green. The local minima of singularity exponents within each positive half-period of the auxiliary functional are taken as GCIs.

properties and does not suffer from usual instabilities. We also present an experiment to show how such sparse solution may result in more realistic estimates of the vocal tract by decoupling of the contributions of the excitation source from that of the vocal tract filter. See figure 11 .

- Multi-pulse estimation of speech excitation source (Vahid Khanagha, Khalid Daoudi). In the GCI detector algorithm, the cardinality of MSM was restricted to one sample per pitch period. We then proceed to study the significance of MSMs of higher cardinalities, in the framework of multi-pulse estimation of voiced sound excitation source. Multi-pulse source coding has been widely used and studied within the framework of Linear Predictive Coding (LPC). It consists in finding a sparse representation of the excitation source (or residual) which yields a source-filter reconstruction with high perceptual quality. The MultiPulse Excitation (MPE) method is the first and one of the most popular techniques to achieve this goal. MPE provides a sparse excitation sequence through an iterative Analysis-by-Synthesis procedure to find the position and amplitudes of the excitation source in two stages: first the location of pulses are estimated one at a time by minimization of perceptually wieghted reconstruction error. In the second stage, the amplitude of these pulses are jointly re-optimized to find the optimal pulse values. Using the MSM, we propose a novel approach to find the locations of the multi-pulse sequence that approximates the speech source excitation. We consider locations of MSM points as the locations of excitation impulses and then, the amplitude of these impulses are computed using the second stage of the classical MPE coder by minimization of the spectrally weighted mean squared error of reconstruction. The multi pulse sequence is then fed to the classical LPC synthesizer to reconstruct speech. Our algorithm is more efficient than classical methods, while providing the same level of perceptual quality as the classical MPE method. See figure 12 .
- Speech representation based local singularity analysis (Vahid Khanagha, Khalid Daoudi, Hussein Yahia, Oriol Pont). Precise estimation of singularity exponents unlocks the determination a collection of points inside the complex signal which are considered as the least predictable points (the MSM). This leads to the associated compact representation and reconstruction. This work presents the very first steps in establishing the links between the MSM and the speech signal. To do so, we make slight modifications to the formalism so as to adapt it to the particularities of the speech signal. Indeed,

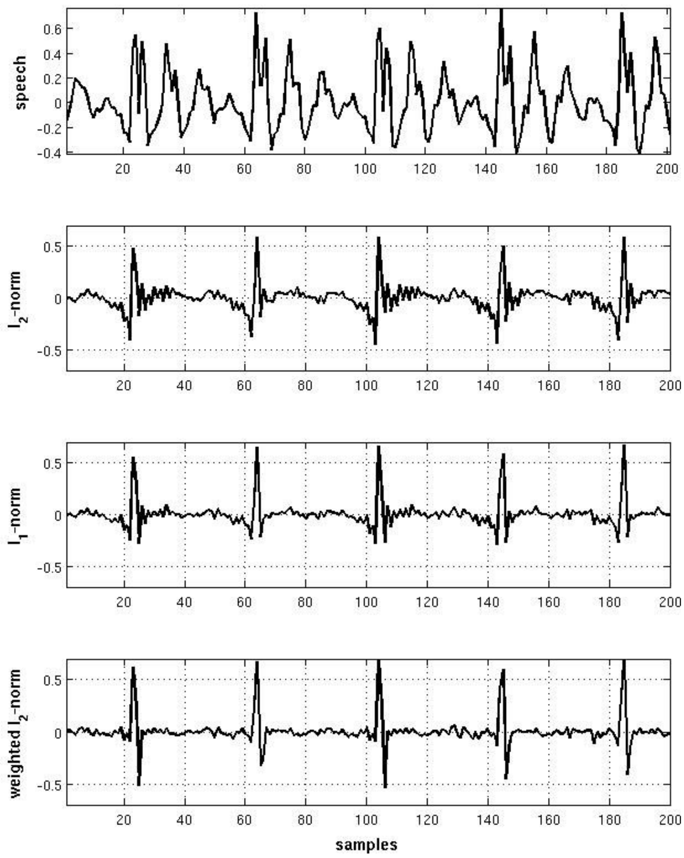


Figure 11. The residuals of the LP analysis obtained from different optimization strategies.

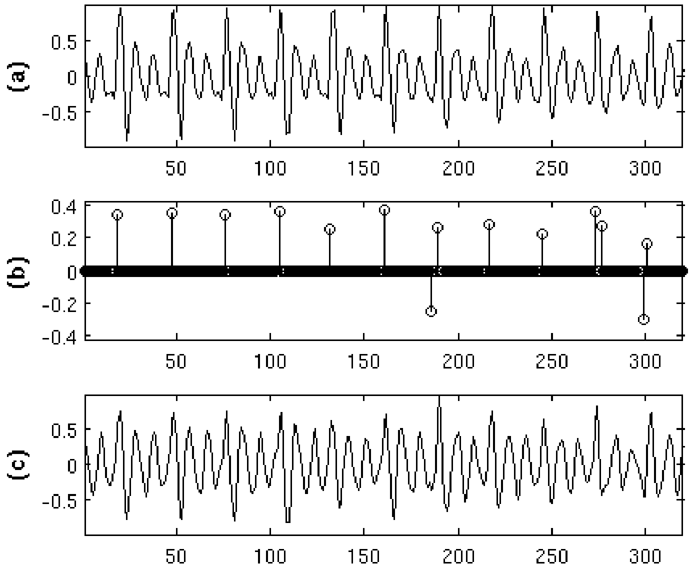


Figure 12. (a) a 40 ms segment of stationary voiced speech, (b) the MSM excitation sequence using 7 pulses per 20 ms and (c) the corresponding reconstructed signal.

the complex intertwining of different dynamics in speech (added to purely turbulent descriptions) suggests the definition of appropriate multi-scale functionals that might influence the evaluation of SEs, hence resulting in a more parsimonious MSM. We present a study that comforts these observations: we show that an alternative multi-scale functional does lead to a more parsimonious MSM from which the whole speech signal can be reconstructed with good perceptual quality. As MSM is composed of a collection of irregularly spaced samples, we use a classical method for the interpolation of irregularly spaced samples, called the Sauer-Allebach algorithm, to reconstruct the speech signal from its MSM. We show that by using this generic algorithm [and even by slight violation of its conditions] high quality speech reconstruction can still be achieved from a MSM of low cardinality. This shows that the MSM formed using the new multi-scale functional we define, indeed can give access to a subset of potentially interesting points in the domain of speech signal. Finally, in order to show the potential of this parsimonious representation in practical speech processing applications, we quantize and encode the MSM so as to develop a waveform coder. See figure 13 .

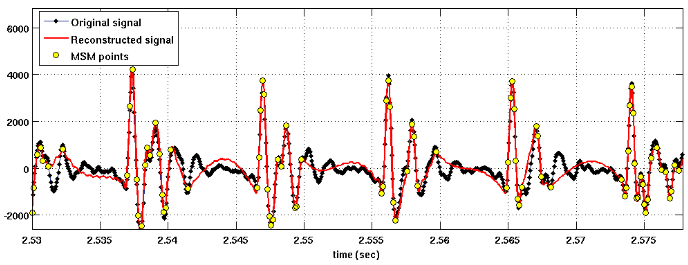


Figure 13. Waveforms of the original signal and the reconstructed signal. Samples belonging to MSM are marked with yellow circles.

- Related publications: [10], [17], [11].

6.8. Reconstruction and gradient-based video editing

Participants: Hicham Badri [correspondant], Hussein Yahia, Driss Aboutajdine.

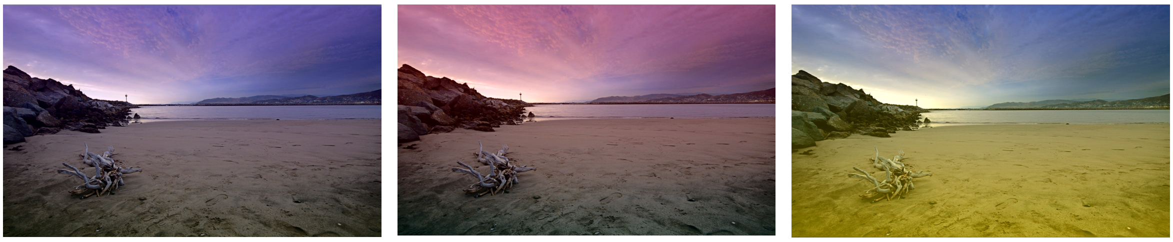


Figure 14. From left to right: original image and examples of non-photoralistic rendering.

Figure 15. Top left: original image. Top right: object removal with FFT-reconstruction algorithm. Botton left: object removal with MVC (Mean Value Coordinates) algorithm. Bottom right: object removal by numerical solving of Poisson equation.

Gradient-domain methods have become a standard for many computational photography applications including object cloning, panorama stitching and non-photorealistic rendering. Integration from a vector field is required to perform gradient-domain-based applications and this operation must be fast enough for interactive editing. The most popular way to perform this integration is known as the Poisson equation and requires solving a large linear system that becomes more costly as the region of interest becomes larger. We propose to use an FFT-based solution and the framework of reconstructible systems instead of performing interactive local/global editing in the gradient domain on the CPU/GPU for both images and videos. See figures 14 , 15 .

- Related publication: [21].

I4S Team

6. New Results

6.1. identification of linear systems

6.1.1. Modular identification and damage detection for large structures

Participants: Michael Döhler, Laurent Mevel.

Subspace identification algorithms are efficient for output-only eigenstructure identification of linear MIMO systems. The problem of merging sensor data obtained from moving and nonsimultaneously recorded measurement setups under varying excitation is considered. To address the problem of dimension explosion, when retrieving the system matrices of the complete system, a modular and scalable approach is proposed. Adapted to a large class of subspace methods, observability matrices are normalized and merged to retrieve global system matrices [12].

6.1.2. Fast multi order subspace identification algorithm

Participants: Michael Döhler, Laurent Mevel.

Subspace methods have proven to be efficient for the identification of linear time-invariant systems, especially applied to mechanical, civil or aeronautical structures in operation conditions. Therein, system identification results are needed at multiple (over-specified) model orders in order to distinguish the true structural modes from spurious modes using the so-called stabilization diagrams. In this paper, new efficient algorithms are derived for this multi-order system identification with subspace-based identification algorithms and the closely related Eigensystem Realization Algorithm. It is shown that the new algorithms are significantly faster than the conventional algorithms in use. They are demonstrated on the system identification of a large-scale civil structure [11], [15].

6.1.3. Evaluation of confidence intervals and computation of sensitivities for subspace methods

Participants: Michael Döhler, Laurent Mevel.

In Operational Modal Analysis, the modal parameters (natural frequencies, damping ratios and mode shapes) obtained from Stochastic Subspace Identification (SSI) of a structure, are afflicted with statistical uncertainty. Uncertainty computation schemes have been developed.This approach has been validated on large scale examples[16].

6.1.4. Subspace methods in frequency domain

Participants: Philippe Mellinger, Michael Döhler, Laurent Mevel.

In Operational Modal Analysis (OMA) of large structures it is often needed to process output-only sensor data from multiple non-simultaneously recorded measurement setups, where some reference sensors stay fixed, while the others are moved between the setups. A standard approach to process the data together for global system identification is to transfer the data into frequency domain and merge it there. However, this only works if the unmeasured ambient excitation remains stationary throughout all setups. As the ambient excitation can be different from setup to setup, the amplitude of the measured data can be different as well and the data has to be normalized. Recently, a method has been developed for covariance- and data-driven Stochastic Subspace Identification (SSI) to automatically normalize and merge the data from multiple setups in order to obtain the global modal parameters (natural frequencies, damping ratios, mode shapes), instead of doing the SSI for each setup separately. In this paper, we adapt this approach to multi-setup SSI in frequency domain, where we use spectra data instead of time series data. We demonstrate the advantages of the new merging approach in the frequency domain and apply it to a relevant industrial large scale example, where we compare the estimation results of the modal parameters between the time and frequency domain approaches [24].

6.1.5. Subspace Identification for Linear Periodically Time-varying Systems

Participant: Ahmed Jhinaoui.

In this paper, an extension of the output-only subspace identification, to the class of linear periodically time-varying (LPTV) systems, is proposed. The goal is to identify a useful information about the system's stability using the Floquet theory which gives a necessary and sufficient condition for stability analysis. This information is retrieved from a matrix called the monodromy matrix, which is extracted by some simultaneous singular value decomposition (SVD) and from a resolution of a least squares criterion. The method is, finally, illustrated by a simulation of a model of a helicopter with hinged-blades rotor and a prototype of the same model. The method is then applied to data from a real wind turbine [22], [19], [20].

6.2. damage detection for mechanical structures

6.2.1. Damage detection and localisation

Participants: Michael Döhler, Luciano Marin, Laurent Mevel.

Mechanical systems under vibration excitation are prime candidate for being modeled by linear time invariant systems. Damage detection in such systems relates to the monitoring of the changes in the eigenstructure of the corresponding linear system, and thus reflects changes in modal parameters (frequencies, damping, mode shapes) and finally in the finite element model of the structure. Damage localization using both finite element information and modal parameters estimated from ambient vibration data collected from sensors is possible by the Stochastic Dynamic Damage Location Vector (SDDLV) approach. Damage is related to some residual derived from the kernel of the difference between transfer matrices in both reference and damage states and a model of the reference state. Deciding that this residual is zero is up to now done using some empirically defined threshold. In this paper, we show how the derivation of the uncertainty of the state space system can be used to derive uncertainty on the damage localization residuals and help to decide about the damage location [23].

6.2.2. Robust subspace damage detection

Participants: Michael Döhler, Laurent Mevel.

Subspace methods enjoy some popularity, especially in mechanical engineering, where large model orders have to be considered. In the context of detecting changes in the structural properties and the modal parameters linked to them, some subspace based fault detection residual has been recently proposed and applied successfully. However, most works assume that the unmeasured ambient excitation level during measurements of the structure in the reference and possibly damaged condition stays constant, which is not possible in any application. This work addresses the problem of robustness of such fault detection methods. A subspace-based fault detection test is derived that is robust to excitation change but also to numerical instabilities that could arise easily in the computations [17], [26].

6.2.3. Input-Output Subspace-Based Fault Detection

Participant: Laurent Mevel.

Subspace-based fault detection method using input-output information is developed in this paper. In some practical applications, the environment noise is the only input that excites the system. Although the statistical properties of the noise might be estimated, the value of the noise is not usually available at each time instance. The traditional subspace fault detection is already developed for such situations. In several other applications, measured inputs are applied to the system or even the stochastic noise might be measurable. While it is still possible to use the traditional output-only detection method, it is reasonable to expect that the application of extra input information together with the output data improves the detection. Several computation issues are discussed in this paper to include input data in the detection method, correctly. Simulation results show the efficiency of using the input information to improve the quality of fault detection [18].

6.2.4. Structural Reliability Updating with Stochastic Subspace Damage Detection Information

Participant: Michael Döhler.

Damage detection algorithms as a part of Structural Health Monitoring (SHM) are widely applied in research and industry and have shown their capabilities to efficiently detect structural damages. These algorithms usually compare a model from a safe reference state of a structure to vibration data from a possibly damaged state. For such a comparison, special properties of real vibration data introduce uncertainties, such as low signal-to-noise ratios, non-stationary or nonwhite ambient excitation, non-linear behavior and many more. Recently, statistical damage detection algorithms based on stochastic subspace identification have been proposed that take into account the uncertainties in the data. Building upon the uncertainty modeling, the next step in the view of the authors is to utilize damage detection algorithm information in the context of the structural reliability theory. Therefore, this paper introduces an approach for the updating of the structural reliability with damage detection algorithm information. Two steps are described namely the determination of a probability of detection (PoD) distribution function for damage detection algorithms accounting for the relevant uncertainties and the concept of Bayesian updating of the structural reliability. The introduced approaches are applied in generic examples. In this way the potential of the utilization of damage detection system information for more reliable structural systems are demonstrated [27].

6.3. Instability monitoring of aeronautical structures

6.3.1. Instability monitoring for LPTV systems

Participants: Laurent Mevel, Ahmed Jhinaoui.

Most subspace-based methods enabling instability monitoring are restricted to the linear time-invariant (LTI) systems. In this paper, a new subspace method of instability monitoring is proposed for the linear periodically time-varying (LPTV) case. For some LPTV systems, the system transition matrices may depend on some parameter and are also periodic in time. A certain range of values for the parameter leads to an unstable transition matrix. Early warning should be given when the system gets close to that region, taking into account the time variation of the system. Using the theory of Floquet, some symptom parameter of stability- or residual- is defined. Then, the parameter variation is tracked by performing a set of parallel cumulative sum (CUSUM) tests. Finally, the method is tested on a simulated model of a helicopter with hinged blades, for monitoring the ground resonance phenomenon [21]. It follows the work on linear systems for aircraft monitoring done previously [14].

6.3.2. Optimal input design for identification and detection

Participant: Laurent Mevel.

This paper considers the problem of auxiliary input design for subspace-based fault detection methods. In several real applications, particularly in the damage detection of mechanical structures and vibrating systems, environment noise is the only input to the system. In some applications, white noise produces low quality output data for the subspace-based fault detection method. In those methods, a residual is calculated to detect the fault based on the output information. However, some modes of the system may not influence the outputs and the residual appropriately if the input is not exciting enough for those modes. In this paper, the method of "rotated inputs" is implemented to excite the system modes. In addition to produce a residual more sensitive to the weak modes, it is possible to detect system order changes due to the fault using the rotated inputs. Simulation results demonstrate the efficiency of injecting the auxiliary input to improve the subspace-based fault detection methodology. [13]. This work is funded by FP7-NMP Large Scale Integrated Project IRIS.

IPSO Project-Team

5. New Results

5.1. PIROCK: a swiss-knife partitioned implicit-explicit orthogonal Runge-Kutta Chebyshev integrator for stiff diffusion-advection-reaction problems with or without noise

In [37], a partitioned implicit-explicit orthogonal Runge-Kutta method (PIROCK) is proposed for the time integration of diffusion-advection-reaction problems with possibly severely stiff reaction terms and stiff stochastic terms. The diffusion terms are solved by the explicit second order orthogonal Chebyshev method (ROCK2), while the stiff reaction terms (solved implicitly) and the advection and noise terms (solved explicitly) are integrated in the algorithm as finishing procedures. It is shown that the various coupling (between diffusion, reaction, advection and noise) can be stabilized in the PIROCK method. The method, implemented in a single black-box code that is fully adaptive, provides error estimators for the various terms present in the problem, and requires from the user solely the right-hand side of the differential equation. Numerical experiments and comparisons with existing Chebyshev methods, IMEX methods and partitioned methods show the efficiency and flexibility of our new algorithm.

5.2. Mean-square A-stable diagonally drift-implicit integrators of weak second order for stiff Itô stochastic differential equations

In [38], we introduce two drift-diagonally-implicit and derivative-free integrators for stiff systems of Itô stochastic differential equations with general non-commutative noise which have weak order 2 and deterministic order 2, 3, respectively. The methods are shown to be mean-square A-stable for the usual complex scalar linear test problem with multiplicative noise and improve significantly the stability properties of the drift-diagonally-implicit methods previously introduced [K. Debrabant and A. Röß ler, Appl. Num. Math., 59, 2009].

5.3. Weak second order explicit stabilized methods for stiff stochastic differential equations

In [39], we introduce a new family of explicit integrators for stiff Itô stochastic differential equations (SDEs) of weak order two. These numerical methods belong to the class of one-step stabilized methods with extended stability domains and do not suffer from the stepsize reduction faced by standard explicit methods. The family is based on the standard second order orthogonal Runge-Kutta Chebyshev methods (ROCK2) for deterministic problems. The convergence, and the mean-square and asymptotic stability properties of the methods are analyzed. Numerical experiments, including applications to nonlinear SDEs and parabolic stochastic partial differential equations are presented and confirm the theoretical results.

5.4. High weak order methods for stochastic differential equations based on modified equations

Inspired by recent advances in the theory of modified differential equations, we propose in [11], a new methodology for constructing numerical integrators with high weak order for the time integration of stochastic differential equations. This approach is illustrated with the constructions of new methods of weak order two, in particular, semi-implicit integrators well suited for stiff (mean-square stable) stochastic problems, and implicit integrators that exactly conserve all quadratic first integrals of a stochastic dynamical system. Numerical examples confirm the theoretical results and show the versatility of our methodology.

5.5. Analysis of the finite element heterogeneous multiscale method for nonmonotone elliptic homogenization problems

In [13], an analysis of the finite element heterogeneous multiscale method for a class of quasilinear elliptic homogenization problems of nonmonotone type is proposed. We obtain optimal convergence results for dimension $d \leq 3$. Our results, which also take into account the microscale discretization, are valid for both simplicial and quadrilateral finite elements. Optimal a-priori error estimates are obtained for the H^1 and L^2 norms, error bounds similar as for linear elliptic problems are derived for the resonance error. Uniqueness of a numerical solution is proved. Moreover, the Newton method used to compute the solution is shown to converge. Numerical experiments confirm the theoretical convergence rates and illustrate the behavior of the numerical method for various nonlinear problems.

5.6. Coupling heterogeneous multiscale FEM with Runge-Kutta methods for parabolic homogenization problems: a fully discrete space-time analysis

Numerical methods for parabolic homogenization problems combining finite element methods (FEMs) in space with Runge-Kutta methods in time are proposed in [14]. The space discretization is based on the coupling of macro and micro finite element methods following the framework of the Heterogeneous Multiscale Method (HMM). We present a fully-discrete analysis in both space and time. Our analysis relies on new (optimal) error bounds in the norms $L^2(H^1)$, $C^0(L^2)$, and $C^0(H^1)$ for the fully discrete analysis in space. These bounds can then be used to derive fully discrete space-time error estimates for a variety of Runge-Kutta methods, including implicit methods (e.g., Radau methods) and explicit stabilized method (e.g., Chebyshev methods). Numerical experiments confirm our theoretical convergence rates and illustrate the performance of the methods.

5.7. A priori error estimates for finite element methods with numerical quadrature for nonmonotone nonlinear elliptic problems

The effect of numerical quadrature in finite element methods for solving quasilinear elliptic problems of nonmonotone type is studied in [12]. Under similar assumption on the quadrature formula as for linear problems, optimal error estimates in the L^2 and the H^1 norms are proved. The numerical solution obtained from the finite element method with quadrature formula is shown to be unique for a sufficiently fine mesh. The analysis is valid for both simplicial and rectangular finite elements of arbitrary order. Numerical experiments corroborate the theoretical convergence rates.

5.8. An Isogeometric Analysis Approach for the study of the gyrokinetic quasi-neutrality equation

In [25], a new discretization scheme of the gyrokinetic quasi-neutrality equation is proposed. It is based on Isogeometric Analysis; the IGA which relies on NURBS functions, seems to accommodate arbitrary coordinates and the use of complicated computation domains. Moreover, arbitrary high order degree of basis functions can be used. Here, this approach is successfully tested on elliptic problems like the quasi-neutrality equation.

5.9. Guiding-center simulations on curvilinear meshes using semi-Lagrangian conservative methods

The purpose of this work [32] is to design simulation tools for magnetised plasmas in the ITER project framework. The specic issue we consider is the simulation of turbulent transport in the core of a Tokamak plasma, for which a 5D gyrokinetic model is generally used, where the fast gyromotion of the particles in the strong magnetic field is averaged in order to remove the associated fast time-scale and to reduce the dimension of 6D phase space involved in the full Vlasov model. Very accurate schemes and efficient parallel algorithms are required to cope with these still very costly simulations. The presence of a strong magnetic field constrains the time scales of the particle motion along and accross the magnetic field line, the latter being at least an order of magnitude slower. This also has an impact on the spatial variations of the observables. Therefore, the efficiency of the algorithm can be improved considerably by aligning the mesh with the magnetic field lines. For this reason, we study the behavior of semi-Lagrangian solvers in curvilinear coordinates. Before tackling the full gyrokinetic model in a future work, we consider here the reduced 2D Guiding-Center model. We introduce our numerical algorithm and provide some numerical results showing its good properties.

5.10. Quasi-periodic solutions of the 2D Euler equation

In [45], we consider the two-dimensional Euler equation with periodic boundary conditions. We construct time quasi-periodic solutions of this equation made of localized travelling profiles with compact support propagating over a stationary state depending on only one variable. The direction of propagation is orthogonal to this variable, and the support is concentrated on flat strips of the stationary state. The frequencies of the solution are given by the locally constant velocities associated with the stationary state.

5.11. Kinetic/fluid micro-macro numerical schemes for Vlasov-Poisson-BGK equation using particles

This work [24] is devoted to the numerical simulation of the Vlasov equation in the fluid limit using particles. To that purpose, we first perform a micro-macro decomposition as in [Benoune, Lemou, Mieussens, JCP 08] where asymptotic preserving schemes have been derived in the fluid limit. In [Benoune, Lemou, Mieussens, JCP 08] , a uniform grid was used to approximate both the micro and the macro part of the full distribution function. Here, we modify this approach by using a particle approximation for the kinetic (micro) part, the fluid (macro) part being always discretized by standard finite volume schemes. There are many advantages in doing so: (i) the so-obtained scheme presents a much less level of noise compared to the standard particle method; (ii) the computational cost of the micro-macro model is reduced in the fluid regime since a small number of particles is needed for the micro part; (iii) the scheme is asymptotic preserving in the sense that it is consistent with the kinetic equation in the rarefied regime and it degenerates into a uniformly (with respect to the Knudsen number) consistent (and deterministic) approximation of the limiting equation in the fluid regime.

5.12. Two-Scale Macro-Micro decomposition of the Vlasov equation with a strong magnetic field

In this paper [26], we build a Two-Scale Macro-Micro decomposition of the Vlasov equation with a strong magnetic field. This consists in writing the solution of this equation as a sum of two oscillating functions with circonscribed oscillations. The first of these functions has a shape which is close to the shape of the Two-Scale limit of the solution and the second one is a correction built to offset this imposed shape. The aim of such a decomposition is to be the starting point for the construction of Two-Scale Asymptotic-Preserving Schemes.

5.13. A dynamic multi-scale model for transient radiative transfer calculations

In [55], a dynamic multi-scale model which couples the transient radiative transfer equation (RTE) and the diffusion equation (DE) is proposed and validated. It is based on a domain decomposition method where the system is divided into a mesoscopic subdomain, where the RTE is solved, and a macroscopic subdomain where the DE is solved. A buffer zone is introduced between the mesoscopic and the macroscopic subdomains, as proposed by [Degond, Jin, SIAM J. Num. Anal. 05], where a coupled system of two equations, one at the mesoscopic and the other at the macroscopic scale, is solved. The DE and the RTE are coupled through the equations inside the buffer zone, instead of being coupled through a geometric interface like in standard domain decomposition methods. One main advantage is that no boundary or interface conditions are needed for the DE. The model is compared to Monte Carlo, finite volume and P1 solutions in one dimensional stationary and transient test cases, and presents promising results in terms of trade-off between accuracy and computational requirements.

5.14. Accuracy of unperturbed motion of particles in a gyrokinetic semi-Lagrangian code

Inaccurate description of the equilibrium can yield to spurious effects in gyrokinetic turbulence simulations. Also, the Vlasov solver and time integration schemes impact the conservation of physical quantities, especially in long-term simulations. Equilibrium and Vlasov solver have to be tuned in order to preserve constant states (equilibrium) and to provide good conservation property along time (mass to begin with). Several illustrative simple test cases are given in [36] to show typical spurious effects that one can observes for poor settings. We explain why Forward Semi-Lagrangian scheme bring us some benefits. Some toroidal and cylindrical GYSELA runs are shown that use FSL.

5.15. High order Runge-Kutta-Nyström splitting methods for the Vlasov-Poisson equation

In this work [46], we derive the order conditions for fourth order time splitting schemes in the case of the $1D$ Vlasov-Poisson system. Computations to obtain such conditions are motivated by the specific Poisson structure of the Vlasov-Poisson system : this structure is similar to Runge-Kutta-Nyström systems. The obtained conditions are proved to be the same as RKN conditions derived for ODE up to the fourth order. Numerical results are performed and show the benefit of using high order splitting schemes in that context.

5.16. A Discontinuous Galerkin semi-Lagrangian solver for the guiding-center problem

In this paper [49], we test an innovative numerical scheme for the simulation of the guiding-center model, of interest in the domain of plasma physics, namely for fusion devices. We propose a 1D Discontinuous Galerkin (DG) discretization, whose basis are the Lagrange polynomials interpolating the Gauss points inside each cell, coupled to a conservative semi-Lagrangian (SL) strategy. Then, we pass to the 2D setting by means of a second-order Strangsplitting strategy. In order to solve the 2D Poisson equation on the DG discretization, we adapt the spectral strategy used for equally-spaced meshes to our Gauss-point-based basis. The 1D solver is validated on a standard benchmark for the nonlinear advection; then, the 2D solver is tested against the swirling deformation ow test case; nally, we pass to the simulation of the guiding-center model, and compare our numerical results to those given by the Backward Semi-Lagrangian method.

5.17. Asymptotic preserving schemes for highly oscillatory kinetic equation

This work [48] is devoted to the numerical simulation of a Vlasov-Poisson model describing a charged particle beam under the action of a rapidly oscillating external electric field. We construct an Asymptotic Preserving numerical scheme for this kinetic equation in the highly oscillatory limit. This scheme enables to simulate the problem without using any time step refinement technique. Moreover, since our numerical method is not based on the derivation of the simulation of asymptotic models, it works in the regime where the solution does not oscillate rapidly, and in the highly oscillatory regime as well. Our method is based on a "double-scale" reformulation of the initial equation, with the introduction of an additional periodic variable.

5.18. Asymptotic preserving schemes for the Wigner-Poisson-BGK equations in the diffusion limit

This work [47] focusses on the numerical simulation of the Wigner-Poisson-BGK equation in the diffusion asymptotics. Our strategy is based on a ”micro-macro" decomposition, which leads to a system of equations that couple the macroscopic evolution (diffusion) to a microscopic kinetic contribution for the fluctuations. A semi-implicit discretization provides a numerical scheme which is stable with respect to the small parameter ε (mean free path) and which possesses the following properties: (i) it enjoys the asymptotic preserving property in the diffusive limit; (ii) it recovers a standard discretization of the Wigner-Poisson equation in the collisionless regime. Numerical experiments confirm the good behaviour of the numerical scheme in both regimes. The case of a spatially dependent $\varepsilon(x)$ is also investigated.

5.19. Orbital stability of spherical galactic models

In [31], we consider the three dimensional gravitational Vlasov Poisson system which is a canonical model in astrophysics to describe the dynamics of galactic clusters. A well known conjecture (Binney, Tremaine in Galactic Dynamics, Princeton University Press, Princeton, 1987) is the stability of spherical models which are nonincreasing radially symmetric steady states solutions. This conjecture was proved at the linear level by several authors in the continuation of the breakthrough work by Antonov (Sov. Astron. 4:859-867, 1961). In the previous work (Lemou et al. in A new variational approach to the stability of gravitational systems, submitted, 2011), we derived the stability of anisotropic models under spherically symmetric perturbations using fundamental monotonicity properties of the Hamiltonian under suitable generalized symmetric rearrangements first observed in the physics literature (Lynden-Bell in Mon. Not. R. Astron. Soc. 144:189-217, 1969; Gardner in Phys. Fluids 6:839-840, 1963; Wiechen et al. in Mon. Not. R. Astron. Soc. 223:623-646, 1988; Aly in Mon. Not. R. Astron. Soc. 241:15, 1989). In this work, we show how this approach combined with a new generalized Antonov type coercivity property implies the orbital stability of spherical models under general perturbations.

5.20. Stable ground states and self-similar blow-up solutions for the gravitational Vlasov-Manev system

In this work [54], we study the orbital stability of steady states and the existence of blow-up self-similar solutions to the so-called Vlasov-Manev (VM) system. This system is a kinetic model which has a similar Vlasov structure as the classical Vlasov-Poisson system, but is coupled to a potential in $-1/r - 1/r^2$ (Manev potential) instead of the usual gravitational potential in -1/r, and in particular the potential field does not satisfy a Poisson equation but a fractional- Laplacian equation. We first prove the orbital stability of the ground states type solutions which are constructed as minimizers of the Hamiltonian, following the classical strategy: compactness of the minimizing sequences and the rigidity of the flow. However, in driving this analysis, there are two mathematical obstacles: the first one is related to the possible blow-up of solutions to the VM system, which we overcome by imposing a sub-critical condition on the constraints of the variational problem. The second difficulty (and the most important) is related to the nature of the Euleri-Lagrange equations (fractional-Laplacian equations) to which classical results for the Poisson equation do not extend. We overcome this difficulty by proving the uniqueness of the minimizer under equimeasurabilty constraints, using only the regularity of the potential and not the fractional- Laplacian Euler-Lagrange equations itself. In the second part of this work, we prove the existence of exact self-similar blow-up solutions to the Vlasov-Manev equation, with initial data arbitrarily close to ground states. This construction is based on a suitable variational problem with equimeasurability constraint.

5.21. Micro-macro schemes for kinetic equations including boundary layers

In this paper [53], we introduce a new micro-macro decomposition of collisional kinetic equations in the specific case of the diffusion limit, which naturally incorporates the incoming boundary conditions. The idea is to write the distribution function f in all its domain as the sum of an equilibrium adapted to the boundary (which is not the usual equilibrium associated with f) and a remaining kinetic part. This equilibrium is defined such that its incoming velocity moments coincide with the incoming velocity moments of the distribution function. A consequence of this strategy is that no artificial boundary condition is needed in the micromacro models and the exact boundary condition on f is naturally transposed to the macro part of the model. This method provides an “Asymptotic preserving" numerical scheme which generates a very good approximation of the space boundary values at the diffusive limit, without any mesh refinement in the boundary layers. Our numerical results are in very good agreement with the exact so-called Chandrasekhar value, which is explicitely known in some simple cases.

5.22. Stroboscopic averaging for the nonlinear Schrödinger equation

In this paper [35], we are concerned with an averaging procedure, -namely Stroboscopic averaging-, for highly-oscillatory evolution equations posed in a (possibly infinite dimensional) Banach space, typically partial differential equations (PDEs) in a high-frequency regime where only one frequency is present. We construct a high order averaged system whose solution remains exponentially close to the exact one over long time intervals, possesses the same geometric properties (structure, invariants, . . .) as compared to the original system, and is non-oscillatory. We then apply our results to the nonlinear Schrödinger equation on the d-dimensional torus T^d, or in R^d with a harmonic oscillator, for which we obtain a hierarchy of Hamiltonian averaged models. Our results are illustrated numerically on several examples borrowed from the recent literature.

5.23. An asymptotic preserving scheme based on a new formulation for NLS in the semiclassical limit

In [41], we consider the semiclassical limit for the nonlinear Schrödinger equation. We introduce a phase/amplitude representation given by a system similar to the hydrodynamical formulation, whose novelty consists in including some asymptotically vanishing viscosity. We prove that the system is always locally well-posed in a class of Sobolev spaces, and globally well-posed for a fixed positive Planck constant in the one-dimensional case. We propose a second order numerical scheme which is asymptotic preserving. Before singularities appear in the limiting Euler equation, we recover the quadratic physical observables as well as the wave function with mesh size and time step independent of the Planck constant. This approach is also well suited to the linear Schrödinger equation.

5.24. Analysis of a large number of Markov chains competing for transitions

In [17], we consider the behaviour of a stochastic system composed of several identically distributed, but non independent, discrete-time absorbing Markov chains competing at each instant for a transition. The competition consists in determining at each instant, using a given probability distribution, the only Markov chain allowed to make a transition. We analyse the first time at which one of the Markov chains reaches its absorbing state. When the number of Markov chains goes to infinity, we analyse the asymptotic behaviour of the system for an arbitrary probability mass function governing the competition. We give conditions that ensure the existence of the asymptotic distribution and we show how these results apply to cluster-based distributed storage when the competition is handled using a geometric distribution.

5.25. High frequency behavior of the Maxwell-Bloch mdel with relaxations: convergence to the Schrödinger-rate system

We study in [20] the Maxwell-Bloch model, which describes the propagation of a laser through a material and the associated interaction between laser and matter (polarization of the atoms through light propagation, photon emission and absorption, etc.). The laser field is described through Maxwell's equations, a classical equation, while matter is represented at a quantum level and satisfies a quantum Liouville equation known as the Bloch model. Coupling between laser and matter is described through a quadratic source term in both equations. The model also takes into account partial relaxation effects, namely the trend of matter to return to its natural thermodynamic equilibrium. The whole system involves 6+$N\,(N+1)/2$ unknowns, the six-dimensional electromagnetic field plus the $N\,(N+1)/2$ unknowns describing the state of matter, where N is the number of atomic energy levels of the considered material. We consider at once a high-frequency and weak coupling situation, in the general case of anisotropic electromagnetic fields that are subject to diffraction. Degenerate energy levels are allowed. The whole system is stiff and involves strong nonlinearities. We show the convergence to a nonstiff, nonlinear, coupled Schrödinger-Boltzmann model, involving 3+N unknowns. The electromagnetic field is eventually described through its envelope, one unknown vector in C^3. It satisfies a Schrödinger equation that takes into account propagation and diffraction of light inside the material. Matter on the other hand is described through a N-dimensional vector describing the occupation numbers of each atomic level. It satisfies a Boltzmann equation that describes the jumps of the electrons between the various atomic energy levels, as induced by the interaction with light. The rate of exchange between the atomic levels is proportional to the intensity of the laser field. The whole system is the physically natural nonlinear model. In order to provide an important and explicit example, we completely analyze the specific (two dimensional) Transverse Magnetic case, for which formulae turn out to be simpler. Technically speaking, our analysis does not enter the usual mathematical framework of geometric optics: it is more singular, and requires an *ad hoc* Ansatz.

5.26. Radiation condition at infinity for the high-frequency Helmholtz equation: optimality of a non-refocusing criterion

In [43], we consider the high frequency Helmholtz equation with a variable refraction index $n^2(x)$ ($x \in \mathbb{R}^d$), supplemented with a given high frequency source term supported near the origin $x = 0$. A small absorption parameter $\alpha_\varepsilon > 0$ is added, which prescribes a radiation condition at infinity for the considered Helmholtz equation. The semi-classical parameter is $\varepsilon > 0$. We let ε and α_ε go to zero *simultaneously*. We study the question whether the prescribed radiation condition at infinity is satisfied *uniformly* along the asymptotic process $\varepsilon \to 0$. This question has been previously studied by the first author, who has proved that the radiation condition is indeed satisfied uniformly in ε, provided the refraction index satisfies a specific *non-refocusing condition*. The non-refocusing condition requires, in essence, that the rays of geometric optics naturally associated with the high-frequency Helmholtz operator, and that are sent from the origin $x = 0$ at time $t = 0$, should not refocus at some later time $t > 0$ near the origin again. In the present text we show the *optimality* of the above mentioned non-refocusing condition. We exhibit a refraction index which *does* refocus the rays of geometric optics sent from the origin near the origin again, and we show that the limiting solution *does not* satisfy the natural radiation condition at infinity in that case.

5.27. Coexistence phenomena and global bifurcation structure in a chemostat-like model with species-dependent diffusion rates

We study in [44] the competition of two species for a single resource in a chemostat. In the simplest space-homogeneous situation, it is known that only one species survives, namely the best competitor. In order to exhibit *coexistence* phenomena, where the two competitors are able to survive, we consider a space dependent situation: we assume that the two species and the resource follow a diffusion process in space, on top of the competition process. Besides, and in order to consider the most general case, we assume each population is associated with a *distinct* diffusion constant. This is a key difficulty in our analysis: the specific (and classical) case where all diffusion constants are equal, leads to a particular conservation law, which in turn allows to eliminate the resource in the equations, a fact that considerably simplifies the analysis and the qualitative phenomena. Using the global bifurcation theory, we prove that the underlying 2-species, stationary, diffusive, chemostat-like model, does possess *coexistence solutions*, where both species survive. On top of that, we identify the domain, in the space of the identified bifurcation parameters, for which the system does have coexistence solutions.

5.28. Markov Chains Competing for Transitions: Application to Large-Scale Distributed Systems

In [16], we consider the behaviour of a stochastic system composed of several identically distributed, but non independent, discrete-time absorbing Markov chains competing at each instant for a transition. The competition consists in determining at each instant, using a given probability distribution, the only Markov chain allowed to make a transition. We analyse the first time at which one of the Markov chains reaches its absorbing state. When the number of Markov chains goes to infinity, we analyse the asymptotic behaviour of the system for an arbitrary probability mass function governing the competition. We give conditions that ensure the existence of the asymptotic distribution and we show how these results apply to cluster-based distributed storage when the competition is handled using a geometric distribution.

5.29. Optimized high-order splitting methods for some classes of parabolic equations

In this paper [21], we are concernedwith the numerical solution obtained by splitting methods of certain parabolic partial differential equations. Splitting schemes of order higher than two with real coefficients necessarily involve negative coefficients. It has been demonstrated that this second-order barrier can be overcome by using splitting methods with complex-valued coefficients (with positive real parts). In this way, methods of orders 3 to 14 by using the Suzuki-Yoshida triple (and quadruple) jump composition procedure have been explicitly built. Here we reconsider this technique and show that it is inherently bounded to order 14 and clearly sub-optimal with respect to error constants. As an alternative, we solve directly the algebraic equations arising from the order conditions and construct methods of orders 6 and 8 that are the most accurate ones available at present time, even when low accuracies are desired. We also show that, in the general case, 14 is not an order barrier for splitting methods with complex coefficients with positive real part by building explicitly a method of order 16 as a composition of methods of order 8.

5.30. A formal series approach to averaging: exponentially small error estimates

The techniques, based on formal series and combinatorics, used nowadays to analyze numerical integrators may be applied to perform high-order averaging in oscillatory periodic or quasi-periodic dynamical systems. When this approach is employed, the averaged system may be written in terms of (i) scalar coefficients that are universal, i.e. independent of the system under consideration and (ii) basis functions that may be written in an explicit, systematic way in terms of the derivatives of the Fourier coefficients of the vector field being averaged. The coefficients may be recursively computed in a simple fashion. We show in [22] that this approach may be used to obtain exponentially small error estimates, as those first derived by Neishtadt. All the constants that feature in the estimates have a simple explicit expression.

5.31. Higher-order averaging, formal series and numerical integration II: the quasi-periodic case

The paper [23] considers non-autonomous oscillatory systems of ordinary differential equations with $d>1$ non-resonant constant frequencies. Formal series like those used nowadays to analyze the properties of numerical integrators are employed to construct higher-order averaged systems and the required changes of variables. With the new approach, the averaged system and the change of variables consist of vector-valued functions that may be written down immediately and scalar coefficients that are universal in the sense that they do not depend on the specific system being averaged and may therefore be computed once and for all. The new method may be applied to obtain a variety of averaged systems. In particular we study the quasi-stroboscopic averaged system characterized by the property that the true oscillatory solution and the averaged solution coincide at the initial time. We show that quasi- stroboscopic averaging is a geometric procedure because it is independent of the particular choice of co-ordinates used to write the given system. As a consequence, quasi-stroboscopic averaging of a canonical Hamiltonian (resp. of a divergence-free) system results in a canonical (resp. in a divergence-free) averaged system. We also study the averaging of a family of near-integrable systems where our approach may be used to construct explicitly d formal first integrals for both the given system and its quasi-stroboscopic averaged version. As an application we construct three first integrals of a system that arises as a nonlinear perturbation of five coupled harmonic oscillators with one slow frequency and four resonant fast frequencies.

5.32. Existence of densities for the 3D Navier-Stokes equations driven by Gaussian noise

We prove in [50] three results on the existence of densities for the laws of finite dimensional functionals of the solutions of the stochastic Navier-Stokes equations in dimension 3. In particular, under very mild assumptions on the noise, we prove that finite dimensional projections of the solutions have densities with respect to the Lebesgue measure which have some smoothness when measured in a Besov space. This is proved thanks to a new argument inspired by an idea introduced in Fournier and Printems (2010).

5.33. Diffusion limit for a stochastic kinetic problem

We study in [30] the limit of a kinetic evolution equation involving a small parameter and perturbed by a smooth random term which also involves the small parameter. Generalizing the classical method of perturbed test functions, we show the convergence to the solution of a stochastic diffusion equation.

5.34. Global Existence and Regularity for the 3D Stochastic Primitive Equations of the Ocean and Atmosphere with Multiplicative White Noise

The Primitive Equations are a basic model in the study of large scale Oceanic and Atmospheric dynamics. These systems form the analytical core of the most advanced General Circulation Models. For this reason and due to their challenging nonlinear and anisotropic structure the Primitive Equations have recently received considerable attention from the mathematical community. In view of the complex multi-scale nature of the earth's climate system, many uncertainties appear that should be accounted for in the basic dynamical models of atmospheric and oceanic processes. In the climate community stochastic methods have come into extensive use in this connection. For this reason there has appeared a need to further develop the foundations of nonlinear stochastic partial differential equations in connection with the Primitive Equations and more generally. In this work [29] we study a stochastic version of the Primitive Equations. We establish the global existence of strong, pathwise solutions for these equations in dimension 3 for the case of a nonlinear multiplicative noise. The proof makes use of anisotropic estimates, $L^p\_tL^q\_x$ estimates on the pressure and stopping time arguments.

5.35. Weak backward error analysis for SDEs

We consider in [28] numerical approximations of stochastic differential equations by the Euler method. In the case where the SDE is elliptic or hypoelliptic, we show a weak backward error analysis result in the sense that the generator associated with the numerical solution coincides with the solution of a modified Kolmogorov equation up to high order terms with respect to the stepsize. This implies that every invariant measure of the numerical scheme is close to a modified invariant measure obtained by asymptotic expansion. Moreover, we prove that, up to negligible terms, the dynamic associated with the Euler scheme is exponentially mixing.

5.36. Convergence of stochastic gene networks to hybrid piecewise deterministic processes

In [27], we study the asymptotic behavior of multiscale stochastic gene networks using weak limits of Markov jump processes. Depending on the time and concentration scales of the system we distinguish four types of limits: continuous piecewise deterministic processes (PDP) with switching, PDP with jumps in the continuous variables, averaged PDP, and PDP with singular switching. We justify rigorously the convergence for the four types of limits. The convergence results can be used to simplify the stochastic dynamics of gene network models arising in molecular biology.

5.37. Exponential mixing of the 3D stochastic Navier-Stokes equations driven by mildly degenerate noises

In [15], we prove the strong Feller property and exponential mixing for 3D stochastic Navier-Stokes equation driven by mildly degenerate noises (i.e. all but finitely many Fourier modes are forced) via Kolmogorov equation approach.

5.38. Existence and stability of solitons for fully discrete approximations of the nonlinear Schrödinger equation

In [40] we study the long time behavior of a discrete approximation in time and space of the cubic nonlinear Schrödinger equation on the real line. More precisely, we consider a symplectic time splitting integrator applied to a discrete nonlinear Schrödinger equation with additional Dirichlet boundary conditions on a large interval. We give conditions ensuring the existence of a numerical soliton which is close in energy norm to the continuous soliton. Such result is valid under a CFL condition between the time and space stepsizes. Furthermore we prove that if the initial datum is symmetric and close to the continuous soliton, then the associated numerical solution remains close to the orbit of the continuous soliton for very long times.

5.39. Fast Weak-Kam Integrators

We consider in [42] a numerical scheme for Hamilton-Jacobi equations based on a direct discretization of the Lax-Oleinik semi-group. We prove that this method is convergent with respect to the time and space stepsizes provided the solution is Lipschitz, and give an error estimate. Moreover, we prove that the numerical scheme is a *geometric integrator* satisfying a discrete weak-KAM theorem which allows to control its long time behavior. Taking advantage of a fast algorithm for computing min-plus convolutions based on the decomposition of the function into concave and convex parts, we show that the numerical scheme can be implemented in a very efficient way.

5.40. Sparse spectral approximations for computing polynomial functionals

In [51], we give a new fast method for evaluating spectral approximations of nonlinear polynomial functionals. We prove that the new algorithm is convergent if the functions considered are smooth enough, under a general assumption on the spectral eigenfunctions that turns out to be satisfied in many cases, including the Fourier and Hermite basis.

MATHRISK Team

5. New Results

5.1. Dynamic risk measures and BSDEs with jumps

The standard approach of mathematical quantification of financial risk in terms of Value at Risk has serious deficiencies. This has motivated a systematic analysis of risk measures which satisfy some minimal requirements of coherence and consistency. The theory of risk measures has been first developed in [54] in the coherent case and then extended in various directions (convex, dynamic, law-invariant) (see e.g. [70], [68], [93], [69]). We are extending this theory, in particular in the case of markets with possible random jumps and model ambiguity, and investigate various types of optimization problems involving risk measures.

Mathematical techniques for the treatment of such problems are based on non linear expectations, backward stochastic differential equations (BSDEs), stochastic control, stochastic differential games.

In the Brownian case, links between dynamic risk measures and Backward Stochastic Differential Equations (BSDEs) have been established (see, among others, [57]). A. Sulem and M.-C. Quenez are exploring these links in the case of stochastic processes with jumps. To this purpose, we have recently extended some comparison theorems for BSDEs with jumps given in [90], and provided a representation theorem of convex dynamic risk measures induced by BSDEs with jumps (see [44]). Optimization of dynamic risk measures leads to stochastic differential games or to optimal control problems for coupled systems of forward-backward stochastic differential equations (FBSDEs). They can be studied by stochastic maximum principles [100] or by transforming them into controlled Backward Stochastic Partial Differential Equations (BSPDEs). We address these questions in collaboration with B. Øksendal (Oslo university) and T. Zhang (Manchester University).

The numerical study of (F)BSDEs with jumps is especially demanding in high dimensions and collaboration has started on these issues with J. Lelong (ENSIMAG) and C. Labart (Université de Savoie).

5.2. Stochastic Differential Games

In many situations, controls are chosen by several agents who interact in various ways. To handle such cases one may use the theory of SDGs. This applies to model uncertainty problems, which can be regarded as a zero-sum game between the agent and the "market" and risk minimization, with risk represented via dynamic risk measures. More general non-zero sum games, involving several players, possibly with asymmetric information or delay will be studied.

An interesting new application of the theory of stochastic differential games, is the issue of *Public Private Partnership* which is a mechanism for a community to outsource the construction of public equipment. The community agrees to pay a rent to the contractor in order to cover the depreciation of the equipment, the maintenance costs and the financial costs. We want to model such partnerships and to compute and compare Nash equilibria and Stackelberg equilibria when the community is the leader. We would also like to investigate whether the community aversion to debt may lead it to enter such a partnership even this is more costly than constructing and managing the equipment by itself.

5.3. Optimal control of Stochastic Partial Differential equations (SPDEs)

SPDEs appear in the modeling of a number of situations: for example, in dynamic pollution models, in financial models involving interest rate derivatives, in systemic risk modeling. The research issues include optimal control of SPDEs and nonlinear filtering theory, stochastic control of forward-backward systems of SPDEs with imperfect and/or asymmetric information, optimal stochastic control of mean-field systems of SPDEs. We have started to study singular control of SPDEs. We plan to give a method for solving optimal control problems for general, possibly non-Markovian systems of FBSDEs by means of BSPDEs with jumps and associated comparison theorems.

5.4. Optimal stopping

Our research on optimal stopping problems covers the analysis of free boundaries in optimal stopping problems for multidimensional stochastic processes with jumps (Thesis of A. Bouselmi, supervised by D. Lamberton). Numerical issues are also be investigated (Monte Carlo methods, quantization methods, methods based on Malliavin calculus). Even in diffusion models, a realistic dividend modeling introduces jumps in the dynamics : at the dividend dates the spot value of the stock undergoes a jump equal to minus the dividend amount. We plan to take into account this feature in optimal stopping problems (Thesis of M. Jeunesse, supervised by B. Jourdain).

The pricing of American options with irregular payoff such as, for instance, binary options, leads to challenging mathematical problems. Some theoretical properties of optimal stopping problems with irregular payoffs have already been obtained. We now plan to focus on the Markovian case by using viscosity solutions and numerical analysis techniques.

In [45], we study optimal stopping problems for (non necessarily) convex dynamic risk measures induced by BSDEs with jumps and establish their connections with *Reflected* BSDEs with jumps. Such problems are related to optimal stopping for non linear expectations, which has been recently studied by [58] in the convex case only. We also address the case of model ambiguity and its relation with mixed control/optimal stopping problems.

5.5. Analysis of stochastic processes with jumps

The use of stochastic processes with jumps in financial modeling has been constantly increasing in the recent years. Simulation of these processes raises specific difficulties. A PhD thesis (V. Rabiet, adviser: V. Bally) has started on regularity properties of the law of multi-dimensional processes with jumps and on sensitivity analysis of derivative products with singular payoffs in such models.

5.6. Monte-Carlo methods

5.6.1. Adaptive variance reduction methods.

Stochastic algorithms [52], [53], [80], [63], [81], [27] or, more recently, direct stochastic optimization [73] proved to be a promising path to automatic variance reduction methods. Direct stochastic optimization techniques are easier to use in practice, avoiding completely any manual tuning needed for stochastic algorithms. This method is well understood (see [73]) only in the Gaussian case and for regular functions. We plan to extend the algorithms and prove rigorous results in non Gaussian cases with financial applications in view for jumps models (see [77], [76], [75]).

5.6.2. Monte-Carlo methods for calibration.

The interest for models combining local and stochastic volatility has been growing recently. Indeed, the local volatility model is not rich enough to efficiently deal with complex derivatives. A popular model is the so called Heston model, in which the volatility process solves a square-root stochastic differential equation (just as in the Cox-Ingersoll-Ross model for interest rate modeling). The thesis of L. Abbas-Turki [12](advisers: D. Lamberton and B. Lapeyre) concentrates on the multi-dimensional Heston model. For these models, numerical aspects are very demanding and we plan to use Monte-Carlo methods using advanced parallel devices (GPU clusters,...) both for price computations and calibration procedures. The thesis of Abbas-Turki is supported by the *Pôle de Compétitivité Finance Innovation* within the consortium *CrediNext*.

5.7. Systemic Risk

We extend the model in [51] in two major ways: First, study the optimal intervention strategy by a lender of last resort that would minimize the size of contagion under budget constraints. Second, allow our model not to be constrained to a single type of financial distress and model jointly insolvency and illiquidity. The interplay of these two mechanisms yields a more potent type of contagion than just the mechanical balance-sheet insolvency type of contagion [85], [92]. In [35], we have started to tackle these issues. This study can be enriched in many different manners.

Benjamin Jourdain and Agnès Sulem have organized a CEA-EDF-Inria school (70 participants) on the issues of Systemic risk and quantitative risk management in October 15-17 2012. (http://bit.ly/finance_inria). A special issue on "Systemic Risk" of the journal *Statistics and Risk Modeling* with B. Jourdain and A. Sulem as guest editors will be published in 2013.

MAXPLUS Project-Team

6. New Results

6.1. Théorie spectrale max-plus et géométrie métrique/Max-plus spectral theory and metric geometry

6.1.1. Introduction

Participants: Marianne Akian, Stéphane Gaubert, Cormac Walsh.

Étant donné un noyau $a : S \times S \to \mathbb{R} \cup \{-\infty\}$, on peut lui associer le problème spectral max-plus

$$\sup_{y \in S} a(x,y) + u(y) = \lambda + u(x), \quad \forall x \in S, \tag{5}$$

dans lequel on cherche le vecteur propre $u : S \to \mathbb{R} \cup \{-\infty\}$ et la valeur propre correspondante $\lambda \in \mathbb{R} \cup \{-\infty\}$. Comme nous l'avons rappelé dans les §3.2 et 3.3 , le problème spectral (9) intervient en contrôle ergodique: l'ensemble S est l'espace des états, et l'application $a(x,y)$ fournit le gain associé à la transition $x \to y$. Le cas où S est fini est classique, l'on a alors un résultat précis de représentation de l'espace propre, à l'aide d'un certain graphe, dit graphe critique. Des résultats existent également lorsque S est compact et que le noyau vérifie certaines propriétés de régularité.

Dans [64], nous avons considéré le cas où S est non compact. Lorsque $\lambda = 0$, l'espace propre est analogue à l'espace des fonctions harmoniques défini en théorie (classique ou probabiliste) du potentiel. En introduisant l'analogue max-plus de la frontière de Martin, nous avons obtenu un analogue de la formule de représentation de Poisson des fonctions harmoniques : toute solution u de (9) peut être représentée sous la forme :

$$u = \sup_{w \in \mathcal{M}_m} w + \mu_u(w) \quad , \tag{6}$$

où $\mathcal{M}_m \subset (\mathbb{R} \cup \{-\infty\})^S$ est l'analogue max-plus de la frontière de Martin minimale (l'ensemble des fonctions harmoniques extrémales normalisées), et où μ_u joue le rôle de la mesure spectrale. Nous avons montré aussi que les éléments de l'espace de Martin minimal peuvent être caractérisés comme les limites de “quasi-géodésiques”. La frontière de Martin max-plus généralise dans une certaine mesure la frontière d'un espace métrique construite à partir des horo-fonctions (fonctions de Busemann généralisées), ou horo-frontière. Ces résultats inspirent les travaux des sections suivantes, qui portent sur des cas remarquables d'espaces métriques (§6.1.4) ou sur des applications en théorie des jeux (§6.1.2).

English version

Let the kernel $a : S \times S \to \mathbb{R} \cup \{-\infty\}$ be given. One may associate the max-plus spectral equation (9), where the eigenvector $u : S \to \mathbb{R} \cup \{-\infty\}$ and the eigenvalue $\lambda \in \mathbb{R} \cup \{-\infty\}$ are unknown. As we recalled in §3.2 and refmonotone, this spectral problem arises in ergodic optimal control: the set S is the *state space*, and the map $a(x,y)$ is the *transition reward*. The case when S is finite is classical, a precise spectral theorem is known, with a characterisation of the eigenspace in terms of a critical graph. Some results have been shown when S is compact, assuming that the kernel a satisfies some regularity properties.

In [64], we considered the case where S is non-compact. When $\lambda = 0$, the eigenspace is analoguous to the set of harmonic functions defined in classical or probabilistic potential theory. By introducing a max-plus analogue of the classical Martin boundary, we obtained an analogue of the Poisson representation of harmonic functions, showing that any solution u of (9) may be represented as in (10) where $\mathcal{M}_m \subset (\mathbb{R} \cup \{-\infty\})^S$ is a max-plus analogue of the minimal Martin boundary (the set of normalised extremal harmonic functions), and μ_u plays the role of the spectral measure. We also showed that the elements of the minimal Martin boundary can be characterised as limits of certain "almost-geodesics". The max-plus Martin boundary generalises to some extent the boundary of metric spaces defined in terms of horofunctions (generalised Busemann functions), or horoboundary. These results have inspired the work of the next sections, which deal either with remarkable examples of metric spaces (§6.1.4) or applications to zero-sum games (§6.1.2).

6.1.2. *Asymptotiques d'itérées d'applications contractantes au sens large et jeux à somme nulle en horizon long/Asymptotics of iterates of nonexpansive mappings and zero-sum games*

Participants: Jérôme Bolte, Stéphane Gaubert, Guillaume Vigeral.

Dans [116], on a établi des résultats de type Denjoy-Wolff pour l'étude asymptotique de la valeur d'un jeu répété, lorsque l'horizon tend vers l'infini. On s'intéresse pour cela plus généralement au "taux de fuite" $\rho(f) = \lim_{k\to\infty} d(x, f^k(x))/k$ où f est une application contractante au sens large pour une "métrique" d sur un espace X. Dans le cas des jeux, X est l'espace des fonctions continues sur l'ensemble des états, f est l'opérateur de Shapley, la métrique d est la norme sup (ou une métrique faible, non-symmétrique, comme $d(x, y) = \max_i (x_i - y_i)$), et $\rho(f)$ représente le maximum du paiement moyen quand l'état initial varie. On a montré, que si l'espace X est de courbure négative en un sens faible (Busemann), alors il existe une horofonction h telle que $h \circ f \geq h + \rho(f)$. Ceci entraîne par exemple, lorsque l'espace d'état est compact, l'existence d'un état dont la valeur croît linéairement avec un taux $\rho(f)$, lorsque l'horizon croît. On a travaillé cette année à la généralisation de ce résultat au temps continu (semigroupes associés à des équations d'Hamilton-Jacobi-Isaacs).

Par ailleurs, dans un travail avec J. Bolte (eprint récent [43]), on s'est intéressé, dans le cas où X est de dimension finie, à l'existence de la limite $\lim_k f^k(x)/k$ (vecteur de paiement moyen). On montre que cette limite existe si l'application f est définissable dans une structure o-minimale. Ceci généralise des résultats de Bewley, Kohlberg, et Neyman, qui montraient que la limite existe si f est semi-algébrique. L'extension au cas o-minimal permet notamment de traiter des opérateurs de type "log-exp" apparaissant en contrôle sensible au risque. Ce travail traite aussi de la question de savoir si un jeu dont les fonctions de paiement et de transition sont définissables dans une structure o-minimale admet un opérateur de Shapley f définissable. Un contre exemple montre que f n'est pas forcément définissable dans la même structure, mais l'on montre qu'il en est ainsi dès que les probabilités de transition ont une structure séparable.

English version

In [116], we established Denjoy-Wolff type results for the asymptotic behaviour of the value of a zero-sum game, when the horizon tends to infinity. To this end, we consider more generally the "escape rate" $\rho(f) = \lim_{k\to\infty} d(x, f^k(x))/k$ where f is a nonexpansive self-map of a "metric" space (X, d). In the case of games, X is the space of continuous functions on the set of states, f is the Shapley operator, and d is the sup-norm (or a weak, non-symmetric, metric like $d(x, y) = \max_i (x_i - y_i)$), and $\rho(f)$ represents the "maximal mean payoff", the maximum being taken over all possible initial states. We showed that if the space X is of nonpositive curvature in a mild sense (Busemann), then, there exists a horofunction h such that $h \circ f \geq h + \rho(f)$. This implies in particular, when the space state is compact, the existence of an initial state from which the value grows linearly with a rate $\rho(f)$, as a function of the horizon. We worked this year on the generalisation of this result to the continuous time case (semigroups associated to Hamilton-Jacovi-Isaacs PDE).

Moreover, in a joint work with J. Bolte (recent eprint [43]), we considered the case in which X is finite dimensional, and studied the existence of the limit $\lim_k f^k(x)/k$ (mean payoff vector). We showed that this limit does exist as soon as the map f is definable in an o-minimal structure. This generalizes results of Bewley, Kohlberg, and Neyman, who showed that this limit exists if f is semi-algebraic. The extension to the case of o-minimal structures allows one in particular to deal with log-exp type operators arising in risk sensitive control. This work also adresses the question of knowing whether a game with definable payment and transition functions has a Shapley operator that is definable in the same structure. We gave a counter example showing that this may not be the case, but showed that the Shapley operator is definable as soon as the transition probabilities have a separable structure.

6.1.3. Isométries de la géométrie de Hilbert/Isometries of the Hilbert geometry

Participants: Cormac Walsh, Bas Lemmens [Kent University, UK].

L'un des intérêts de l'horo-frontière est de renseigner sur le groupe des isométries d'un espace métrique. En effet, ce groupe agit naturellement sur l'horo-frontière, et cette action peut parfois être mieux comprise que l'action du groupe sur l'espace d'origine.

Nous étudions le groupe des isométries pour la métrique de Hilbert. De La Harpe [181] a donné plusieurs conjectures relatives à ce groupe. Nous conjecturons que le groupe des isométries est exactement le groupe des transformations linéaires projectives à moins que le domaine ne soit une coupe d'un cône symmétrique non-Lorentzien. Nous avons démontré précédemment cette conjecture lorsque le domaine est un polytope [135].

Dans le cas général, on peut prouver, en utilisant les horo-fonctions, que si il existe une bijection entre deux cônes homogéne de degré -1, antitone, et d'inverse antitone, ces deux cônes sont symétriques. Nous essayons maintenant de montrer que toute isométrie de Hilbert sur un domaine convexe est la version projective d'un automorphisme linéaire du cône sur le domaine, ou d'une bijection du cône, homogéne de degré -1, qui est antitone et d'inverse antitone. Ce résultat pemettrait de compléter la preuve de la conjecture proposée plus haut.

L'état actuel de l'étude de ce problème est résumé dans un article de Walsh [41] (chapître d'un "handbook on the Hilbert geometry" à paraître).

English version

One use for the horofunction boundary is to study the group of isometries of a metric space. This is because this group has a well defined action on the horoboundary and it is likely that in many cases this action will be easier to understand than the action on the space itself.

We have been investigating the isometries of the Hilbert geometry. De La Harpe [181] has previously made several conjectures about the isometry group of this space. We conjecture that the isometry group is exactly the group of projective linear transformations unless the domain on which the geometry is defined is a cross section of a non-Lorentzian symmetric cone. We have previously proved that this conjecture is true in the case of a polytope domain [135].

In the general case, we can now prove, using horofunctions, that if a bijection between cones is homogeneous of degree -1, order inverting, and has an order inverting inverse, then both cones are symmetric. We are working on showing that every Hilbert isometry on a convex domain arises by considering projectively either a linear automorphism on the cone over the domain, or a homogeneous -1, order inverting bijection on this cone with order inverting inverse. Establishing this result would complete our proof of the above conjecture.

The current state of knowledge about this problem has been summarized in a paper by Walsh [41] that will appear as a chapter in a forthcoming handbook on the Hilbert geometry.

6.1.4. Espace de Teichmüller/Teichmüller space

Participant: Cormac Walsh.

L'espace de Teichmüller d'une surface est un espace métrique composé des structures conformes de cette surface. On peut le voir comme l'ensemble des classes d'équivalence des métriques riemanniennes de cette surface, où deux métriques sont équivalentes si il existe une application conforme homotope à l'identité qui envoie l'une des métriques sur l'autre.

Il existe plusieurs métriques naturelles sur l'espace de Teichmüller. Nous avons travaillé précédemment sur la métrique Lipschitz de Thurston et avons prouvé [174] que l'horo-frontière de cet espace métrique était la frontière de Thurston.

Néanmoins, la métrique la plus utilisées ur l'espace de Teichmüller est la métrique de Teichmüller. L'horo-frontière de cet espace métrique n'est autre que la frontière déja introduite dans la littérature sous le nom de frontière de Gardiner–Masur. Nous étudions cette frontière, en particulier nous donnons explicitement ses points de Busemann [55].

Par la suite, nous avons l'intention d'utiliser cette propriété afin d'étudier les sous-groupes du groupe modulaire, qui est le groupe des isométries de la métrique de Teichmüller.

English version

An interesting metric space is the Teichmüller space of a surface. This is the space of conformal structures on the surface. One may think of it as the space of equivalence classes of Riemannian metrics on the surface, where two such metrics are regarded as being equivalent if there is a conformal map on the surface taking one to the other that is homotopic to the identity.

There are several natural metrics on Teichmüller space. Previously, we have worked with Thurston's stretch metric and have shown [174] that the horofunction boundary with this metric is just the usual Thurston boundary.

However, the most commonly used metric on Teichmüller space is Teichmüller's metric. The horofunction boundary of this metric space turns out to be the same as a previously defined boundary, called the Gardiner–Masur boundary. We have been investigating this boundary. In particular, we have managed to work out explicitly its Busemann points [55].

In future work, we intend to apply this knowledge to study subgroups of the mapping class group, which is the isometry group of the Teichmüller metric.

6.1.5. Consensus non-commutatif et contraction d'opérateurs de Kraus/Noncommutative consensus and contraction of Kraus maps

Participants: Stéphane Gaubert, Zheng Qu.

Dans un travail récent [44], on s'est intéressé à la vitesse de convergence vers l'équilibre d'une itération de la forme $x^{k+1} = T(x^k)$, $x^k \in X$, où T est une application linéaire préservant un cône dans un espace de Banach X, telle que $T(e) = e$, pour un certain vecteur e dans l'interieur du cône. On s'intéresse aussi à l'itération dans l'espace dual, $y^{k+1} = T^*(y^k)$, $y^k \in X^*$, lorsque $\langle y^0, e\rangle = 1$.

Le cas classique est celui où $T(x) = Px$ est un opérateur de Markov. L'itération primale traduit alors la convergence vers le "consensus", et l'itération duale traduit la convergence de la distribution de probabilité en temps k vers l'état stationnaire. Dans ce cas, le taux de contraction (en un coup) $\kappa(P)$ d'une itération primale, pour la semi-norme de Hilbert $\|z\|_H := \max_i z_i - \min_j z_j$, ainsi que le taux de contraction d'une itération duale, pour la métrique en variation totale, coïncident et sont caractérisés par une formule dûe à Doeblin et Dobrushin (coefficient d'ergodicité),

$$\kappa(P) := 1 - \min_{i,j} \sum_{s=1}^{n} \min(P_{is}, P_{js}).$$

On a donné ici une généralisation de cette formule au cas d'opérateurs abstraits, qui s'applique en particulier aux opérateurs de Kraus qui interviennent en information quantique. Ces derniers opérent sur l'espace des matrices symmétriques, et sont de la forme

$$T(x) = \sum_k a_k x a_k^* \qquad \text{avec} \qquad \sum_k a_k a_k^* = I \ .$$

English version

In a recent work [44], we studied the speed of convergence to equilibrium of an iteration of the form $x^{k+1} = T(x^k)$, $x^k \in X$, where T is a linear map preserving a cone in a Banach space X, such that $T(e) = e$, for some vector e in the interior of the cone. We also considered the iteration in the dual space X^*, $y^{k+1} = T^*(y^k)$, $y^k \in X^*$, where $\langle y^0, e\rangle = 1$.

The classical application arises when $T(x) = Px$ is a Markov operator. Then, the primal iteration represents the dynamics of consensus, whereas the dual iteration represents the evolution of the probability distribution as a function of time. Then, the (one-shot) contraction rate $\kappa(P)$ of the primal iteration, with respect to Hilbert's seminorm $\|z\|_H := \max_i z_i - \min_j z_j$, and the contraction rate of the dual iteration, with respect to the total variation metric, coincide, and are characterized by a formula of Doeblin and Dobrushin (ergodicity coefficient),

$$\kappa(P) := 1 - \min_{i,j} \sum_{s=1}^{n} \min\left(P_{is}, P_{js}\right).$$

We gave here a generalization of this formula to an abstract operators on a cone. This covers in particular the Kraus maps arising in quantum information theory. The latter maps act on the space of symmetric matrices. They can be written as

$$T(x) = \sum_k a_k x a_k^* \qquad \text{with} \qquad \sum_k a_k a_k^* = I \ .$$

6.2. Algèbre linéaire max-plus et convexité abstraite/Max-plus linear algebra and abstract convex analysis

6.2.1. Convexité max-plus ou tropicale/Max-plus or tropical convexity

Participants: Xavier Allamigeon, Stéphane Gaubert, Eric Goubault [CEA], Ricardo Katz [Conicet, Argentine].

On étudie les analogues max-plus ou tropicaux des ensembles convexes. Ceux-ci sont utiles en particulier pour représenter de manière effective les ensembles d'états accessibles de systèmes à événements discrets [9], ils sont aussi apparus récemment en géométrie tropicale, dans toute une série de travaux à la suite de Sturmfels et Develin [98]. Les polyèdres max-plus peuvent aussi être vus comme des limites de déformations de polyèdres classiques, sur lesquels ils donnent un éclairage de nature combinatoire. Toutes ces motivations ont inspiré la recherche d'analogues des résultats fondamentaux d'analyse convexe classique: séparation, projection, points extrémaux, à la suite en particulier de [8].

Dans un travail de X. Allamigeon, S. Gaubert, et E. Goubault [68], [16], on a mis en évidence un critère combinatoire pour la caractérisation des sommets des polyèdres tropicalement convexes. Celui-ci s'exprime à l'aide d'hypergraphes orientés, et de leurs composantes fortement connexes. Ce critère possède la propriété d'être vérifiable en un temps presque linéaire en la taille de l'hypergraphe.

On en déduit un analogue tropical de la méthode de la double description [16] (méthode très utilisée sur les polyèdres classiques, et dûe à Motzkin *et al.* [147]). Cet algorithme permet de calculer les sommets d'un polyèdre défini de façon externe (intersection de demi-espaces ou d'hyperplans tropicaux). Grâce au critère combinatoire précédent, l'algorithme améliore de plusieurs ordres de grandeur les techniques connues jusqu'alors. Ceci est confirmé par de nombreuses expérimentations. Ce travail est motivé par des applications à l'analyse statique [67] et aux systèmes à événements discrets [101], dans lesquelles la manipulation de tels polyèdres est le goulot d'étranglement.

Dans un travail de X. Allamigeon, S. Gaubert, et R. Katz [69], on étend le théorème de McMullen au cas tropical: ce dernier caractérise le nombre maximal de points extrêmes d'un polyèdre, en fonction du nombre d'inégalités qui le définissent et de sa dimension. Nous montrons que la même borne est valide dans le cas tropical (à une modification triviale près). Cependant, le calcul de la borne optimale est encore ouvert dans ce cas.

Il est connu qu'un polyèdre tropical peut être représenté comme l'enveloppe convexe d'un ensemble minimal de points et rayons, donnés par ses sommets et ses rayons extrêmes [112]. Dans un travail réalisé par X. Allamigeon et R. Katz [48], et effectué en partie lors d'une visite de R. Katz à Inria (juillet 2011), on étudie la question duale de la caractérisation des représentations minimales par demi-espaces. On montre qu'un polyèdre tropical possède *essentiellement* une unique représentation minimale par demi-espaces, lorsque leurs apex appartiennent au polyèdre. On montre que les apex de ces demi-espaces non-redondants correspondent à certains sommets du complexe tropical introduit par Develin et Sturmfels [98]. On introduit également un critère combinatoire pour l'élimination de demi-espaces redondants à l'aide d'hypergraphes orientés.

Dans un travail en cours de X. Allamigeon, P. Benchimol, S. Gaubert et R. Katz (débuté lors d'une visite de ce dernier à Inria en novembre 2012), nous étudions la tropicalisation des représentations par demi-espaces des polyèdres convexes sur le corps des séries de Puiseux. Nous démontrons ainsi une conjecture de Develin et Yu [99]. Celle-ci assure qu'étant donné un polytope tropical pur, il existe un polytope *lift* sur les séries de Puiseux, dont les demi-espaces associés aux faces se "tropicalisent" en une représentation par demi-espaces du polytope tropical initial.

English version

We study the max-plus or tropical analogues of convex sets. These have been used in particular to represent effectively the accessible sets of certain discrete event systems [9]. They also appeared in tropical geometry, following the work of Sturmfels and Develin [98]. Max-plus polyhedra can be thought of as limits of deformations of classical polyhedra, on which they give a combinatorial insight. These motivations have inspired the investigation of analogues of basic results of classical convex analysis: separation, projection, representation by extreme points, following [8].

In a work of X. Allamigeon, S. Gaubert, and E. Goubault [16], we introduce a combinatorial criterion for the characterization of the vertices of tropically convex polyhedra. It is expressed in terms of directed hypergraphs and their strongly connected components. This criterion can be verified in almost linear time in the size of the hypergraph.

This allows to develop a tropical analogue of the double description method [16] (this method is widely used for classical convex polyhedra, and is due to Motzkin *et al.* [147]). This algorithm is able to determine all the vertices of a polyhedron defined externally (intersection of tropical half-spaces of hyperplanes). Thanks to the combinatorial criterion mentioned above, the algorithm improves the existing methods by several orders of magnitude. This is confirmed by several experiments. This is motivated by applications to static analysis [67] and discrete event systems [101], in which computing such polyhedra turns out to be the bottleneck.

In a work of X. Allamigeon, S. Gaubert, and R. Katz [69], we extend the McMullen upper bound theorem to the tropical case. This theorem characterises the maximal number of extreme points of a polyhedron, as a function of the number of inequalities defining it, and of the dimension. We show that the same bound is valid in the tropical case (up to a trivial modification). However, computing the optimal bound is an open problem in this case.

It is well-known that a tropical polyhedron can be represented as the convex hull of a minimal set of points and rays, provided by its vertices and extreme rays [112]. In a work of X. Allamigeon and R. Katz [48], partly done during the visit of R. Katz at Inria (July 2011), the dual problem of characterizing the minimal representations by half-spaces is studied. We show that a tropical polyhedron admits *essentially* a unique minimal external representation by half-spaces, provided that their apices belong to the polyhedron. We prove that the apices of these half-spaces correspond to certain vertices of the tropical complex introduced by Develin and Sturmfels [98]. We also establish a combinatorial criterion allowing to eliminate redundant half-spaces using directed hypergraphs.

In an ongoing work of X. Allamigeon, P. Benchimol, S. Gaubert and R. Katz (started during a visit of the latter at Inria in Novembre 2012), we study the tropicalization of the representation by half-spaces of convex polyhedra over the field of Puiseux series. In particular, we prove a conjecture of Develin and Yu [99]. It states that, given a pure tropical polytope, there exists a lifting polytope over Puiseux series, such that the facet-defining half-spaces are "tropicalized" into a representation by half-spaces of the initial polytope.

6.2.2. Convexes max-plus et jeux avec paiements ergodiques/Max-plus convex sets and mean payoff games

Participants: Marianne Akian, Xavier Allamigeon, Stéphane Gaubert, Alexander Guterman [Moscow State University], Ricardo Katz [Conicet, Argentine], Sergei Sergeev [Birmingham, UK].

Dans un travail d'Akian, Gaubert et Guterman [15], on a montré un résultat d'équivalence entre les jeux ergodiques à somme nulle et les systèmes d'inégalités max-plus linéaires: décider la non-vacuité d'un polyèdre tropical est équivalent à vérifier si un jeu déterministe à somme nulle a un paiement moyen par unité de temps positif ou nul. Plus généralement, la même question pour un jeu stochastique à somme nulle est équivalente à vérifier si un convexe tropical (non-polyédral, i.e., défini par un système infini d'inégalités) est vide. Ces résultats sont démontrés à l'aide de techniques de théorie de Perron-Frobenius non-linéaire. Ils sont ensuite appliqués à l'étude de l'indépendance linéaire dans le semi-anneau tropical.

Le résultat de [15] a eu plusieurs retombées.

Dans un travail de Gaubert et Sergeev [24], on réduit le problème spectral tropical de type faisceaux, $Ax = \lambda Bx$, à un jeu paramétrique (ce qui permet de calculer le spectre en temps pseudo-polynômial).

Enfin, dans un travail de Gaubert, Katz, et Sergeev [22], on développe un algorithme de programmation linéaire tropicale (pseudo-polynômial) basé sur cette correspondance avec les jeux répétés.

English version

In a work by Akian, Gaubert and Guterman [15], we showed the equivalence mean payoff games and max-plus linear inequalities: testing whether a tropical polyhedron is non-empty is equivalent to checking whether a mean payoff deterministic game is winning. More generally, checking whether a mean payoff stochastic game is winning is equivalent to checking the non-emptyness of a tropical convex set defined by an infinite family of inequalities. These results are established using techniques of non-linear Perron-Frobenius theory. Then, they are applied to the study of linear independence over the tropical semiring.

The equivalence established in [15] had several consequences.

In a work of Gaubert and Sergeev [24], the tropical spectral problem for matrix pencils, $Ax = \lambda Bx$, is reduced to a parametric game (which allows one to compute the spectrum in pseudo-polynomial time).

Finally, in a work of Gaubert, Katz, and Sergeev [22], a (pseudo-polynomial) tropical linear programming algorithm is developed, based on the same correspondence with mean payoff games.

6.3. Algèbre max-plus, déformations et asymptotiques /Max-plus algebra, deformations and asymptotic analysis

6.3.1. Introduction

Comme indiqué dans le §3.7 , l'algèbre max-plus est la limite d'une déformation de l'algèbre classique, ou plutôt du semi-corps des réels positifs. Elle peut aussi fournir des estimations de ces déformations, puisque

$$\max(a,b) \le \epsilon \log(e^{a/\epsilon} + e^{b/\epsilon}) \le \epsilon \log(2) + \max(a,b) \quad . \tag{7}$$

L'utilisation de ces propriétés a déjà conduit dans le passé aux travaux sur les perturbations de valeurs propres [60], [59], [58], ou sur les grandes déviations [1], [62]. Dans les travaux qui suivent, nous exploitons ces propriétés dans des contextes reliés ou similaires à ceux de nos travaux précédents.

English version

As detailled in §3.7 , max-plus algebra is the limit of a deformation of classical algebra, or more precisely of the semi-field of usual real positive numbers. It can also give estimations for these deformations using for instance (11). By using these properties, we already obtained some works on singular perturbations of matrix eigenvalues [60], [59], [58], or on large deviations [1], [62]. In the works described below, we are exploiting again these properties in contexts that are related or similar to those of our earlier works.

6.3.2. Aspects tropicaux des algorithmes de scaling matriciel/Tropical aspects of matrix scaling problems

Participants: Marianne Akian, Stéphane Gaubert, Meisam Sharify Najafabadi [LRI, Paris Sud].

Une partie du travail de thèse de M. Sharify [167] portait sur les méthodes de mise à l'échelle pour améliorer la précision du calcul de valeurs propres. En appliquant les techniques de [58], [59], on montrait notamment que l'ordre de grandeur des valeurs propres d'un faisceau matriciel est donné (sous des conditions de non-dégénerescence) par les valeurs propres tropicales, qui peuvent être calculées de manière robuste, et fournissent ainsi une mise à l'échelle pour calculer les valeurs propres classiques.

Nous avons poursuivi ce travail dans [47], qui a été présenté dans [36]. On calcule cette fois l'ordre de grandeur des valeurs propres d'un polynôme matriciel au moyen des racines tropicales du polynôme obtenu en appliquant une norme donnée aux coefficients. Les racines dépendent de la norme choisie, et la norme de Frobenius est optimale en un certain sens. On obtient des bornes générales pour les ratios entre modules des valeurs propres et racines tropicales qui généralisent les bornes obtenues par Polya et Ostrowski dans le cas de polynômes scalaires. On raffine aussi ces bornes, en particulier lorsque les racines tropicales sont bien séparées les unes des autres.

English version

A part of the PhD work of M. Sharify [167] dealt with scaling methods to improve the accuracy of eigenvalue numerical computions. Applying the techniques of [58], [59], we shown in particular that the order of magnitude of the eigenvalues of a matrix pencil can be determined (under nondegeracy conditions) by computing tropical eigenvalues. The latter can always be computed accurately and provide a scaling which can be combined with standard numerical methods for matrix pencils.

We have pursued this work in [47], which has been presented in [36]. Now we compute the order of magnitude of the eigenvalues of a matrix polynomial by using the tropical roots of the polynomial the coefficients of which are obtained by applying a norm to the coefficients of the matrix polynomial. The tropical roots depend on the chosen norm, and the Frobenius turns out to be optimal in a certain sense. We obtain indeed general bounds on the ratios between the modulus of the eigenvalues of the matrix polynomial and the tropical roots which generalize the bounds of Polya and Ostrowski available for scalar polynomials. We also improve these bounds, in particular when the tropical roots are well separated.

6.3.3. Méthodes tropicales de localisation de valeurs propres de matrices/Tropical methods for the localisation of matrix eigenvalues

Participants: Marianne Akian, Stéphane Gaubert, Andrea Marchesini.

Lors de son stage de M2 dans l'équipe, Andrea Marchesini a obtenu des inégalités de type majorisation entre les valeurs propres d'une matrice et les valeurs propres tropicales de la matrice de ses modules. En particulier, la majoration est une généralisation de l'inégalité de Friedland [108] concernant le rayon spectral.

La thèse d'Andrea Marchesini s'inscrit dans le prolongement de son stage de M2 dans l'équipe et certains des travaux de la thèse de Meisam Sharify [167]. Le but est d'obtenir des inégalités de type majorisation permettant d'estimer a priori les valeurs propres de matrices ou de faisceaux de matrices, en faisant éventuellement intervenir des hypothèses de bon conditionnements. En particulier on recherche la localisation de ces valeurs propres en fonction de valeurs propres de matrices agrégées ou simplifiées. On cherchera aussi à obtenir le même type de localisation ou d'estimation dans le cas des vecteurs propres associés, par exemple en utilisant les techniques de compléments de Schur de [59] ou les idées de Murota [148].

L'idée est ensuite d'utiliser ces résultats de localisation pour améliorer la précision des algorithmes de calcul numérique de valeurs propres de matrices, en particulier en construisant des changements d'échelle exploitant les calculs tropicaux, à effectuer préalablement à l'appel d'algorithmes classiques comme QZ. Les travaux de Stéphane Gaubert et Meisam Sharify [115] ont montré l'intérêt de cette approche, notamment pour les problèmes de faisceaux quadratiques de valeurs propres issus de systèmes mécaniques pour lesquels on dispose de nombreux exemples pathologiques pour les algorithmes existants.

English version

During his M2 internship in the team, Andrea Marchesini has obtained majorization type inequalities between the eigenvalues of a matrix and the tropical eigenvalues of the matrix obtained by applying the modulus entrywise. In particular, the bound is a generalization of the inequality of Friedland [108] concerning the spectral radius.

The PhD thesis follows his M2 internship and some of the works of Meisam Sharify's PhD thesis [167]. The aim is to obtain majorization type inequalities allowing one to estimate the eigenvalues of matrices or matrix polynomials, using possibly assumptions on condition numbers. In particular, one may look for estimates of these eigenvalues using the eigenvalues of aggregated or simplified matrices. One may also try to find the same type of estimates for the associated eigenvectors, for instance by using techniques of Schur complements from [59] or ideas of Murota [148].

One would like to use these estimation results to improve the accuracy of eigenvalue numerical computions, in particular by constructing scaling methods using tropical techniques, which may be used before calling usual algorithms as QZ. The works of Stéphane Gaubert and Meisam Sharify [115] showed the interest of this approach, in particular for quadratic matrix polynomials issued from mechanical systems for which there exists several pathological examples for existing algorithms.

6.3.4. Mesures et applications maxitives/Maxitive measures and maps

Participants: Marianne Akian, Stéphane Gaubert, Paul Poncet.

La thèse de Paul Poncet [154] concernait essentiellement ce que l'on appelle l'analyse idempotente, c'est-à dire l'étude des espaces fonctionnels ou linéaires de dimension infinie sur l'algèbre tropicale, ou tout autre semi-anneau idempotent. Paul Poncet a développé pour cela un point de vue treillis continus comme dans [1], ou plus généralement domaines. Depuis la soutenance en novembre 2011, plusieurs articles issus du manuscrit de thèse sont en cours de publication ou de soumission, et d'autres travaux pousuivant ceux de la thèse sont en cours avec les membres de l'équipe.

La première partie de la thèse traitait des mesures maxitives, en particulier de l'existence d'une densité cardinale ou d'une densité d'une mesure par rapport à une autre (théorème de Radon-Nikodym), et de la régularité d'une mesure maxitive. Paul Poncet donnait en particulier une caractérisation des mesures maxitives régulières à valeurs dans un domaine, qui raffinait le théorème de décomposition des mesures maxitives établi dans [153]. Ces résultats font maintenant l'objet de [54].

La deuxième partie concernait les convexes dans les semi-treillis ou l'algèbre max-plus, pour lesquels Paul Poncet a pu établir des théorèmes de type Krein-Milman, réciproque de Milman, et représentation de Choquet. Un article concernant le cas des semi-treillis [53] a été soumis.

Enfin la troisième et dernière partie qui traitait des semi-groupes inverses dans une tentative d'unification de l'algèbre usuelle et de l'algèbre tropicale fait l'objet de l'article [31].

On sait que les résultats sur les convexes tropicaux de dimension infinie de [154], qui se déduisent soit des résultats correspondants sur les semi-treillis, soit de résultats de théorie des mesures maxitives de la première partie de la thèse [154], permettent de retrouver partiellement les résultats sur la frontière de Martin max-plus décrits dans la section 6.1.1 . Dans un travail commun nous essayons d'obtenir d'autres applications et extensions du théorème de représentation de Choquet tropical. En particulier on considère le cas d'ensembles ordonnés qui ne sont pas forcément des treillis tels que le cône des matrices symmetriques positives muni de l'ordre de Loewner.

English version

The PhD thesis work of Paul Poncet [154] concerned essentially what is called idempotent analysis, that is the study of infinite dimensional functional or linear spaces over tropical algebra, or any other idempotent semiring. For this aim, Paul Poncet developped the point of view of continuous lattices, as in [1], or more generally of domains. Since the defense of his thesis in November 2011, several papers derived from the thesis manuscript have been submitted and some are published or up to be published. Some other works pursuing the thesis work are done with team members.

The first part of the Paul Poncet's thesis concerned maxitive measures, in particular the existence of a cardinal density of a measure, or that of a density of a measure with respect to another (Radon-Nikodym theorem), and the regularity of a maxitive measure. Paul Poncet gave in particular a characterization of domain valued maxitive measures that are regular, which improved the decomposition theorem of maxitive measures stated in [153]. These results are now gathered in [54].

A second part concerned convex sets in lattices or max-plus algebra, for which Paul Poncet showed results such as a Krein-Milman type theorem, a Milman converse type theorem, and a Choquet representation type theorem. A manuscript concerning the case of semilattices [53] has been submitted.

The third and last part which studied inverse semigroups in an attempt to unify usual and tropical algebras is presented in [31].

We know that the results on infinite dimensional tropical convex sets of [154], which are deduced either from the corresponding results on semilattices, or from the results on maxitive measures of the first part of the thesis manuscript, allow one to recover at least partially the results on max-plus Martin boundaries described in Section 6.1.1 . In a joint work, we try to obtain other applications and extensions of the max-plus Choquet representation theorem. In particular, we consider the case of ordered sets that are not necessarily semilattices, such as the cone of nonnegative symmetric matrices endowed with the Loewner order.

6.4. Algorithmes/Algorithms

6.4.1. Méthodes multigrilles pour le contrôle stochastique et les jeux répétés à somme nulle/Multigrid methods for stochastic control and repeated zero sum games

Participants: Marianne Akian, Sylvie Detournay.

L'algorithme d'itération sur les politiques est bien connu pour résoudre efficacement les équations de la programmation dynamique associées à des problèmes de contrôle stochastique avec critère à horizon infini (Howard) ou ergodique (Howard, et Denardo et Fox). Récemment, il a été généralisé au cas de problèmes de jeux à deux joueurs et somme nulle dégénérés (avec paiements ergodiques et de type "multi-chaîne"), au moyen de techniques d'algèbre max-plus et de théorie du potentiel non linéaire [87]. Chaque itération de base de cet algorithme utilise la résolution d'un système d'équations linéaires dont l'opérateur est monotone, mais dont la taille peut être grande, soit parce qu'il provient d'une discrétisation fine d'une équation aux dérivées partielles, soit parce qu'il est associé à un problème discret de grande taille comme le graphe du Web.

Or, la méthode multigrille est l'une des rares méthodes permettant de résoudre, au moins dans les bons cas, des systèmes linéaires en un temps de l'ordre de la taille du système. De plus, alors que la méthode multigrille classique ne s'applique qu'à des discrétisations d'équations aux dérivées partielles elliptiques, la méthode multigrille algébrique (voir par exemple [164]) peut s'appliquer à tout système linéaire présentant des propriétés de monotonie (principe du maximum ou système avec M-matrice).

L'association entre méthodes multigrilles et itérations sur les politiques avait déjà été utilisée et étudiée dans le cas de problèmes de contrôle stochastique actualisé (voir par exemple [57], [65]), ainsi que dans le cas d'un algorithme d'itération sur les politiques simplifié pour le contrôle ergodique (voir par exemple [5]), mais pour lequel il n'existe pas de preuve de convergence. La méthode multigrille algébrique avait été récemment associée à des méthodes d'apprentissage (voir par exemple [180]). Nous l'avions aussi testée dans le cas de l'itération sur les politiques pour des problèmes de jeux à somme nulle actualisés au cours du stage de Shantanu Gangal en 2007.

La thèse de Sylvie Detournay a eu pour but de développer et d'étudier un algorithme associant une méthode d'itération sur les politiques du type de celle introduite par Cochet-Terrasson et Gaubert dans [87] et une méthode multigrille algèbrique, afin de résoudre des problèmes de jeux à somme nulle dégénérés, éventuellement posés directement sous forme discrète. Au cours de sa thèse, Sylvie Detournay a codé l'ensemble des algorithmes en C, en faisant appel éventuellement à des librairies existantes en particulier les méthodes multigrilles algèbriques d'Yvan Notay. L'ensemble des codes nouveaux est déposé sur le projet “pigames” de la gforge et sera disponible librement.

Une première partie de la thèse [11] qui a été publiée dans [14] concerne le cas non dégénéré (actualisé). Elle comprend en particulier des tests sur des discrétisations d'équations aux dérivées partielles d'Hamilton-Jacobi-Bellman ou d'Isaacs, ou d'inéquations variationnelles.

Le reste de la thèse concerne le cas de problèmes avec critère moyen en temps. Sylvie Detournay a en particulier implémenté et raffiné l'algorithme proposé par Cochet-Terrasson et Gaubert [87], en l'associant soit à des méthodes de résolution exacte de systèmes linéaires, soit à des méthodes multigrilles algébriques, en utilisant aussi des méthodes multigrilles multiplicatives pour le calcul de la mesure invariante de chaînes de Markov irréductibles, comme celles introduites par De Sterck. Ceci a permis l'obtention de résultats numériques dans le cas de discrétisations d'équations d'Isaacs associées à des jeux de poursuite déterministes ou aléatoires. Cela a aussi permis de tester de manière systématique l'algorithme sur des instances aléatoires de jeux de type Richman. Certains de ces résultats, ainsi que la présentation de l'algorithme (de manière plus concrète que dans [87], et avec les détails d'implémentation) et les preuves de sa convergence sont regroupés dans le manuscrit [45] écrit avec Jean Cochet-Terrasson et Stéphane Gaubert.

Ces travaux ont aussi conduit à l'introduction dans [11] d'une nouvelle méthode multigrille multiplicative pour le calcul de la mesure invariante de chaînes de Markov irréductibles, qui consiste en l'application de l'algorithme d'itération sur les politiques combiné aux méthodes multigrilles algèbriques au problème de contrôle optimal (à un joueur) avec critère moyen en temps obtenu par transformation log-exp du système linéaire initial. Cette méthode a été testée et comparée aux méthodes multigrilles multiplicatives existantes.

English version

Policy iteration is a powerful and well known algorithm to solve the dynamic programming equation associated to stochatic control (one player game) problems with infinite horizon criterion (Howard) or ergodic criterion (Howard and Denardo and Fox). It has recently been extended to degenerate two players problems (with ergodic payoff and in “multichain” cases) using ideas from max-plus algebra and nonlinear potential theory [87]. One basic iteration of the algorithm consists in solving a linear system the operator of which is monotone, but with a size which may be large since it comes from the discretization of a partial differential equation or since it is associated to a large size discrete problem arising from instance from the Web graph.

For the solution of large size linear systems, the state of art consists of multigrid methods which are often able to solve systems in linear time. Whereas multigrid methods can only be applied to systems that come from discretizations of elliptic partial differential equations, algebraic multigrid methods (see for instance [164]) can be applied to any linear system with monotonicity properties (discrete maximum principle or system with a M-matrix).

The association of multigrid methods with policy iteration has been used and studied in the case of discounted stochastic control problems (see for instance [57], [65]), or in the case of a simplified policy iteration algorithm for ergodic control (see for instance [5]), but for which no proof of convergence is known. Some recent work combines the algebraic multigrid method with learning methods [180]. We also tested it in the case of policy iterations for discounted zero-sum two-player games, during the internship of Shantanu Gangal in 2007.

The aim of the PhD thesis of Sylvie Detournay was to develop and study an algorithm for degenerate two player games (that may come from a discrete time and finite state space model) combining a policy iteration such as the one introduced in [87] and an algebraic multigrid method (AMG). During her thesis, Sylvie Detournay coded all algorithms in C, using eventually existing librairies in particular the algebraic multigrid libray of Yvan Notay. All new algorithms belong to the gforge project "pigames" and will be distributed openly.

A first part of the thesis manuscript [11], which has published in [14], concerns the nondegenerate (discounted) case. It contains in particular some tests on discretisations of Hamilton-Jacobi-Bellman or Isaacs partial differential equations or variational inequalities.

The rest of the thesis concerns the case of problems with mean-payoff criteria. In particular, Sylvie Detournay has implemented and refined the algorithm proposed by Cochet-Terrasson and Gaubert [87], while associating it either to direct linear solvers, or to the AMG methods already used in the nondegenerate case, and using also multiplicative AMG methods for computing invariant measures of Markov chains, such as the one introduced by De Sterck. This allowed her to obtain numerical results in the case of discretisations of Isaacs equations associated to deterministic or stochastic pursuit games. This also allowed her to test systematically the algorithm on random instances of Richman type games. Some of these results, together with the presentation of the algorithm (in a more practical manner than in [87], with implementation details), and convergence proofs are gathered in the article [45] with Jean Cochet-Terrasson and Stéphane Gaubert.

These works also led to the introduction in [11] of a new multiplicative AMG method for computing invariant measures of irreducible Markov chains. This method consists of the application of the policy iteration algorithm combined with AMG method to the optimal control (or one player) problem with mean-payoff criteria obtained after a log-exp transformation of the initial linear system. It has been tested and compared with previous multiplicative AMG methods.

6.4.2. Algorithmique des polyèdres tropicaux/Algorithmics of tropical polyhedra

Participants: Xavier Allamigeon, Pascal Benchimol, Stéphane Gaubert, Eric Goubault [CEA], Michael Joswig [TU Darmstadt].

X. Allamigeon, S. Gaubert, et E. Goubault, ont développé dans [67], [16] plusieurs algorithmes permettant de manipuler des polyèdres tropicaux. Ceux-ci correspondent aux travaux décrits dans §6.2.1 . Ils permettent notamment de déterminer les sommets et rayons extrêmes d'un polyèdre tropical défini comme intersection de demi-espaces, ou inversement, de calculer une représentation externe à partir d'un ensemble de générateurs. Ces algorithmes sont implémentés la bibliothèque TPLib (voir §5.3).

Dans un travail en cours de X. Allamigeon, P. Benchimol, M. Joswig et S. Gaubert, nous nous intéressons aux problèmes de programmation linéaire tropicale. Nous définissons un analogue tropical de la méthode du simplexe. L'algorithme repose sur une technique de pivotage entièrement combinatoire entre deux points de base, se fondant sur la notion d'hypergraphes tangents.

English version

X. Allamigeon, S. Gaubert, and E. Goubault, have developed in [67], [16] algorithms allowing one to manipulate tropical polyhedra. They correspond to the contributions described in §6.2.1 . In particular, they can be used to determine the vertices and extreme rays of a tropical polyhedron defined as the intersection of half-spaces, or inversely, to compute an external description from a set of generators. These algorithms are implemented in the library TPLib (see §5.3).

In an ongoing work of X. Allamigeon, P. Benchimol, M. Joswig and S. Gaubert, we study the problems of tropical linear programming. We define a tropical analog of simplex algorithm. It relies on a pivoting technique between two basis points, which is entirely combinatorial, and which involves the notion of tangent hypergraphs.

6.4.3. Problèmes d'accessibilité dans les hypergraphes orientés et leur complexité/Reachability problems in directed hypergraphs and their complexity

Participant: Xavier Allamigeon.

Les hypergraphes orientés sont une généralisation des graphes orientés, dans lesquelles chaque arc relie un ensemble de sommets à un autre. Ils jouent un rôle important dans les travaux récents sur la convexité tropicale (voir §6.2.1), puisqu'ils offrent une représentation naturelle des cônes définis sur le sous-semi-anneau booléen $\mathbb{B} = \{-\infty, 0\}$.

Dans un travail de X. Allamigeon [17], on étudie la complexité de problèmes d'accessibilité sur les hypergraphes orientés. Nous introduisons un algorithme de complexité presque linéaire permettant de déterminer les composantes fortement connexes terminales (qui n'accèdent à aucune autre composante si ce n'est elles-mêmes) d'un hypergraphe.

Nous établissons également une borne inférieure sur-linéaire sur la taille de la réduction transitive de la relation d'accessibilité dans les hypergraphes. Cela indique que la relation d'accessibilité dans les hypergraphes orientés est combinatoirement plus complexe que celle des graphes orientés. Cela suggère aussi que des problèmes comme le calcul des composantes fortement connexes est plus difficile sur les hypergraphes que sur les graphes. Nous mettons d'ailleurs en évidence une réduction en temps linéaire du problème du calcul des ensembles minimaux dans une famille d'ensembles donnée, vers le problème du calcul de toutes les composantes fortement connexes d'un hypergraphe. Le problème du calcul des ensembles minimaux a été largement étudié dans la littérature [155], [176], [175], [156], [157], [158], [103], [74], et aucune algorithme en temps linéaire n'est connu à ce jour.

English version

Directed hypergraphs are a generalization of directed graphs, in which the tail and the head of the arcs are sets of vertices. It appears that they play an important role in the recent works on tropical convexity (see §6.2.1), since they offer a natural representation of cones defined over the boolean sub-semiring $\mathbb{B} = \{-\infty, 0\}$.

In a work of X. Allamigeon [17], we study the complexity of reachability problems on directed hypergraphs. We introduce an almost linear-time algorithm allowing to determine the terminal strongly connected components (a component is said to be *terminal* when no other component is reachable from it).

We also establish a super-linear lower bound over the size of the transitive reduction of the reachability relation in directed hypergraphs. This indicates that the reachability relation is combinatorially more complex in directed hypergraphs than in directed graphs. This also suggests that reachability problems such as computing all strongly connected components are likely to be harder in hypergraphs than in graphs. Besides, we show that the minimal set problem can be reduced in linear time to the problem of computing all strongly connected components in hypergraphs. The former problem consists in finding all minimal sets among a given family of sets. It has been well studied in the literature [155], [176], [175], [156], [157], [158], [103], [74], and no linear time algorithm is known.

6.4.4. Approximation max-plus de fonctions valeurs et équations de Riccati généralisées/Max-plus approximation of value functions and generalized Riccati equations

Participants: Stéphane Gaubert, Zheng Qu, Shanjian Tang [Fudan University, Shanghai].

La thèse de Zheng Qu, supervisée par S. Gaubert et S. Tang, porte sur le développement de méthodes tropicales en programmation dynamique approchée.

Les méthodes d'approximation max-plus conduisent à approcher la fonction valeur d'un problème de contrôle ou de jeux par un supremum d'un nombre fini de formes quadratiques, voir notamment [114]. On s'intéresse ici à l'analyse théorique (complexité) ainsi qu'à l'amélioration de ces méthodes. Dans certains cas, ces formes quadratiques sont propagées par des flots d'équations de Riccati généralisées. Afin d'effectuer des analyses d'erreur, on exploite les propriétés de contraction du flot de Riccati pour certaines métriques connues sur le cône des matrices positives, et en particulier pour la métrique de Thompson. Celle-ci n'est rien d'autre que $d_T(A, B) = \|\log \mathrm{spec}\,(A^{-1}B)\|_\infty$, où spec désigne la suite des valeurs propres d'une matrice, et log s'entend composante par composante.

Ceci nous a amené à étudier le problème général du calcul du taux de contraction d'un flot monotone sur un cône, pour la métrique de Thompson. En effet, les propriétés de contraction de l'équation de Riccati standard sont connues (résultats de Bougerol pour la métrique Riemanienne invariante, et de Wojtowski pour la métrique de Thompson), mais les techniques de preuve employées dans ce cadre (semigroupes de matrices symplectiques) ne s'étendent pas aux équations généralisées.

On donne dans [51] une formule explicite générale pour le taux de contraction pour la métrique de Thompson d'un flot monotone, faisant seulement intervenir le générateur du flot et sa dérivée. On a notamment appliqué ce résultat à une équation de Riccati généralisée associé à des problèmes de contrôle stochastique avec critère quadratique, dans lesquels la dynamique comporte un terme bilinéaire en le contrôle et le bruit. On a montré dans ce cas que la métrique de Thompson est la seule métrique de Finsler invariante pour laquelle le flot est nonexpansif, et l'on a caractérisé la constante de contraction locale.

Une application de ces résultats à l'analyse d'une méthode de réduction de la malédiction de la dimension, dûe à McEneaney, a été réalisée récemment par Z. Qu.

English version

The PhD work of Zheng Qu is supervised by S. Gaubert and S. Tang, it aims in particular at developing tropical methods in approximate dynamic programming.

The max-plus methods lead to approach the value function of an optimal control or zero-sum game problem by a supremum of a finite number of quadratic forms, see in particular [114]. We are interested here in the theoretical analysis (complexity) of this class of methods, as well as of their improvement. In certain cases, the quadratic forms are propagated by the flows of generalized Riccati equations. In order to perform an error analysis, we need to use some contraction properties of the Riccati flow, for certain known metrics on the space of positive matrices, like Thompson's metric. The latter is nothing but $d_T(A,B) = \|\log \mathrm{spec}\,(A^{-1}B)\|_\infty$, where spec denotes the sequence of eigenvalues of a matrix, and $\log$ is understood entrywise.

This led us to study the general problem of computing the contraction rate of an order-preserving flow on a cone, with respect to Thompson's metric. Indeed, the contraction properties of the standard Riccati flow are known (theorem of Bougerol for the invariant Riemanian metric, of Wojtowski for the Thompson's metric), but the proof of these properties (based on symplectic semigroups) does not carry over to generalized Riccati equations.

We gave in [51] a general explicit formula for the contraction rate with respect to Thompson's metric of an order-preserving flow, involving only the generator of the flow and its derivative. We applied in particular this result to a generalized Riccati equation, associated to stochastic optimal control problems with a quadratic cost and a bilinear dynamics (presence of a bilinear term between the control and the noise). We showed that in this case, the Thompson's metric is the only invariant Finsler metric in which the generalized Riccati flow is nonexpansive, and we characterized the local contraction rate of this flow.

Z. Qu applied recently these results to the analysis of a method of reduction of the curse of dimensionality, introduced by McEneaney.

6.5. Applications

6.5.1. Introduction

Nous présentons maintenant plusieurs travaux de nature appliquée, touchant à des domaines variés, dans lesquels nous exploitons certaines des techniques mathématiques présentées précédemment, et particulièrement celles qui relèvent de la théorie de Perron-Frobenius non-linéaire et de la convexité tropicale. Ces applications utilisent aussi des techniques d'algèbre linéaire ou d'optimisation convexe.

English version

In this section, we describe several applied works in which we use some of the theoretical tools developed by the team, including non-linear Perron-Frobenius theory and tropical convexity. Some of these applications also make an intensive use of linear algebraic and convex programming methods.

6.5.2. Propriétés des valeurs propres de Perron et de Floquet, et application en chronothérapeutique/Properties of Perron and Floquet eigenvalue, with an application to chronotherapeutics

Participants: Frédérique Billy [Projet BANG, Inria], Jean Clairambault [Projet BANG, Inria], Olivier Fercoq, Stéphane Gaubert, Thomas Lepoutre [Projet BANG puis DRACULA, Inria].

On s'intéresse à des modèles de systèmes dynamiques monotones structurés en âge représentant la croissance de populations de cellules (saines ou tumorales), à la suite de travaux de Clairambault et Perthame. Il s'agit de comprendre l'influence du contrôle circardien sur la croissance des cellules. Dans le cas stationnaire, le taux de croissance est représenté par une valeur propre de Perron. Dans le cas périodique, il s'agit d'une valeur propre de Floquet. Les travaux [40], [18], [77] portent sur l'identification de ces modèles ainsi que sur un problème de contrôle thérapeutique, consistant à minimiser le taux de croissance des cellules tumorales sous une contrainte de non-toxicité du traitement (maintien d'une population de cellules saines). Ce travail s'appuie en particulier sur un algorithme d'optimisation de la valeur propre de Perron d'une matrice développé par Fercoq dans un autre contexte [106].

English version

We study monotone dynamical systems representing the growth of cells (healthy or tumoral), following a work of Clairambault and Perthame. The goal is to understand how the circadian control influences the growth of cells. In the case of stationnary monotone systems, this growth is measured by the Perron root. In the time periodic case, this Perron root is replaced by a Floquet multiplier.

The works [40], [18], [77] deal with the identification of these models, together with a therapeutic control problem, consisting in minimizing the growth rate of tumoral cells, under a non-toxicity constraint (preserving the population of healthy cells). This works relies in particular on a fast algorithm to optimize the Perron eigenvalue of a matrix, developed by Fercoq in a different context [106].

6.5.3. Équations aux dérivées partielles en dynamique des populations/Partial differential equations arising in population dynamics

Participants: Sepideh Mirrahimi, Stéphane Gaubert.

Nous étudions la limite en temps long de dynamiques des populations structurées. Il s'agit de l'étude asymptotique de l'équation suivante

$$\partial_t n_\varepsilon - \varepsilon \Delta n_\varepsilon = \frac{n_\varepsilon}{\varepsilon} R(x, I_\varepsilon), \quad I_\varepsilon(t) := \int \psi(x) n_\varepsilon(x,t) dx. \tag{8}$$

Il est connu qu'asymptotiquement, lorsque le taux de mutation est petit et en temps long, la solution de cette équation se concentre en une masse de Dirac en un point de maximum de $R(\cdot, I_M)$, avec $I_M = \lim_{t\to\infty} \lim_{\varepsilon\to 0} I(t)$. Un tel point s'appelle ESS (Evolutionary stable strategy) en dynamiques adaptatives. On s'intéresse à savoir, dans le cas où le problème admet plusieurs ESS (qui correspondent à des points de maximum de R), vers quel ESS la densité va converger en temps grand. Nous essayons de répondre à cette question en supposant que le taux de mutations est important (comme dans le cas des cellules cancéreuses). Nous voudrions déterminer la limite suivante: $\lim_{\varepsilon\to 0} \lim_{t\to\infty} n_\varepsilon(x,t)$. Une conjecture est que la limite est une masse de Dirac en un point x_M où x_M est le point de maximum de $R(\cdot, I_M)$ au voisinage duquel $R(x, I_M)$ est plus plat (une fonction F dépendant de la hessienne de R est maximisée en x_M). Celle-ci est motivée d'une part par un travail de M. Akian, R. Bapat et S. Gaubert, montrant à l'aide d'outils de théorie spectrale max-plus qu'une propriété analogue est vraie en dimension finie (convergence du vecteur propre de Perron de matrices dont les coefficients sont de la forme $\exp(A_{ij}/\varepsilon)$), et d'autre part par des travaux reliés en théorie de KAM faible (les points de maximum de R correspondent à l'ensemble d'Aubry projeté). L'objectif est donc ici de déterminer quel vecteur propre du problème ergodique est sélectionné à la limite visqueuse.

Nous avons déjà identifié la limite lorsque le taux de mutations tend vers 0 ($\varepsilon \to 0$) en partant de la solution stationnaire de (12). Il nous reste à démontrer que la solution de (12) converge en temps long vers la solution stationnaire. L'analogue discret de ce problème est également une question ouverte à laquelle on s'intéresse.

English version

We study the long-time asymptotic behaviour of structured population models. We consider specially the PDE (12). It is known that asymptotically, when the mutation rate is small, and the time horizon is large, the solution of this equation concentrates to a Dirac mass at a maximum point of $R(\cdot, I_M)$, with $I_M = \lim_{t\to\infty} \lim_{\varepsilon\to 0} I(t)$. Such a limit point is called ESS (Evolutionary stable strategy) in the field of adaptative dynamics. We are interested to know, when there are several ESS (corresponding to several points of maximum of R), to which ESS the density will converge as the horizon tends to infinity. We are studying this question in particular when the mutation rate is large (as in the case of tumor cells), leading to compute the following limit: $\lim_{\varepsilon\to 0} \lim_{t\to\infty} n_\varepsilon(x, t)$. We made a conjecture that the limit is a Dirac mass at a point x_M where among the points of maximum of $R(\cdot, I_M)$, x_M is the one at which $R(x, I_M)$ is the "flatest" (an auxiliary function F depending on the Hessian of R is maximized at point x_M). This is motivated on the one hand by a previous work of M. Akian, R. Bapat and S. Gaubert, showing, through max-plus spectral theory, that an analogous property does hold in finite dimension (convergence of the Perron eigenvector of matrices with coefficients $\exp(A_{ij}/\varepsilon)$), and on the other hand, by related works in weak KAM theory (the points of maximum of R correspond to the projected Aubry set); these works determine the eigenvector of the ergodic problem which is selected by the viscous limit.

We already identified the limit when the mutation rate tends to 0 ($\varepsilon \to 0$), starting from the stationnary solution of (12). We still need to show that the solution of (12) does converge in large time to the stationnary solution. Even the discrete analogue of this problem is an open issue, which we are studying.

6.5.4. Analyse statique de programmes et itération sur les politiques/Static analysis of computer programs and policy iteration

Participants: Assale Adjé [LSV, ENS Cachan], Stéphane Gaubert, Eric Goubault [CEA].

On applique ici des méthodes de théorie des jeux et d'optimisation (analyse convexe abstraite, programmation convexe et non convexe) aux problèmes de point fixe intervenant en analyse statique de programme. On a introduit dans [13] un nouveau domaine en analyse statique, qui étend au cas non-linéaire le domaine des "gabarits" introduit par Manna, Sankaranarayanan, and Sipma [166]. Ce domaine permet de représenter des ensembles accessibles non-convexes (définis par un nombre fini d'inégalités prises dans un dictionnaire). Ceci permet d'intégrer en particulier des informations liées à l'existence de fonctions de Lyapunov, qui sont souvent connues dans les applications issues de l'ingénierie. Nous avons montré dans [13] que des invariants (expérimentalement précis) pouvaient être obtenus en couplant l'itération sur les politiques avec des relaxations de Shor (relaxations SDP de problèmes quadratiques non-convexes), ce qui fournit des abstractions précises de certains programmes numériques (ex: filtres avec seuils).

Un problème important consiste à déterminer le plus petit point fixe (l'algorithme de [13] fournit un point fixe, qui peut ne pas être minimal). Ce problème est abordé dans [26], où l'approche de [13] est comparée avec une approche duale développée par Gawlitza et Seidl.

English version

We apply methods from game theory and optimization (generalized duality, convex and non convex programming) to the fixed point problems arising in static analysis of programs by abstract interpretation. We introduced in [13] a new domain in static analysis, which extends to nonlinear cases the "templates" introduced by Manna, Sankaranarayanan, and Sipma [166]. This domain allows one to represent accessible sets that are non convex. These are defined by finitely many inequalities taken from a dictionnary. This allows one to use in particular the information provided by Lyapunov functions, which are often known in applications arising from engineering. We showed in [13] that experimentally accurate invariants can be obtained by coupling policy iteration with Shor relaxation (SDP relaxation of convex programming problems). This yields accurate abstractions of some numerical programs, like linear filters with thresholds.

An important problem consists in determining the smallest fixed point (the algorithm of [13] yields a possibly non minimal fixed point). This problem is addressed in [26], in which the approach of [13] is compared with a dual approach developed by Gawlitza and Seidl.

6.5.5. Optimisation du référencement sur la toile/Optimization of web referencing

Participants: Marianne Akian, Mustapha Bouhtou [Orange Labs], Olivier Fercoq, Stéphane Gaubert.

La thèse d'O. Fercoq [12], co-encadrée par M. Akian, M. Bouhtou, et S. Gaubert, financée par un CRE d'Orange Labs, avait pour but d'appliquer des méthodes d'optimisation et de théorie des jeux à l'optimisation de services en lignes. On a tout d'abord étudié le problème de l'optimisation du référencement, que l'on formalise en se donnant par exemple un ensemble d'hyperliens et de ressources obligatoires, dont la nature et la position sur le site web sont déterminées à l'avance par le concepteur. Cet ensemble forme en quelque sorte le squelette du site web. On se donne aussi un ensemble d'hyperliens ou de ressources facultatives, pour lesquels le concepteur du site a certains degrés de liberté (le lien ou le contenu peut être mis sur une page plutôt qu'une autre, voire être omis).

Dans [20], on aborde le problème de l'optimisation du "Pagerank" dans ce cadre, en appliquant des techniques de décision Markovienne classiques et sous-contraintes. Le problème peut en effet se ramener à un problème de contrôle ergodique ou de contrôle ergodique sous contraintes (ergodiques), selon que les contraintes sur les hyperliens sont locales à chaque page ou font intervenir plusieurs pages. On traite à la fois le cas relaxé où les probabilités de passage d'une page à une autre peuvent être des réels positifs quelconques (on peut par exemple supposer que cette probabilité dépend de la position et des caractères utilisés pour l'hyperlien correspondant) et le cas discret où ces probabilités sont uniformes parmis celles qui sont strictement positives (comme dans la modélisation classique conduisant au calcul du Pagerank). On montre que cette famille de problèmes correspondent à des problèmes de programmation dynamique avec un nombre exponentiel de contrôles, mais où les polytopes des mesures de probabilités de transition admettent des oracles de séparation polynômiaux. On obtient de la sorte des résultats de complexité, ainsi que, sous certaines hypothèses, des algorithmes adaptés à des instances de grande taille, couplant programmation dynamique et relaxation Lagrangienne. Ces algorithmes ont été testés sur un fragment du graphe du web.

Un critère de référencement classique, alternatif au pagerank, est donné par le vecteur propre de Perron, comme dans le cas de l'algorithme "HITS" de Kleinberg. O. Fercoq a abordé le problème associé d'optimisation du référencement, qui se révèle plus difficile que celui du pagerank, en raison de l'absence de propriété de convexité. Cependant, il a développé un algorithme rapide et creux (basé sur des propriétés de rang 1 d'opérateurs intervenant dans le calcul de dérivées du critère) permettant de calculer un optimum local du référencement [106].

O. Fercoq a aussi donné un algorithme analogue pour optimiser le score "HOTS" de Tomlin [38]. Cependant, la convergence de l'algorithme original de HOTS n'avait jamais été prouvée. Dans [50], O. Fercoq a identifié le taux de convergence de l'algorithme et de plusieurs de ses variantes grâce à des techniques d'applications contractantes au sens large et aux propriétés des problèmes de flot d'entropie maximale dans un réseau.

La thèse de Fercoq comprend aussi un algorithme de classement permettant de déterminer les pages de Spam parmi un ensemble de pages douteuses, supposant connues un autre ensemble de pages repertoriées comme spam [33]. Cet algorithme exploite les techniques développées pour l'optimisation du PageRank [33].

English version

The goal of the PhD work [12] of O. Fercoq, cosupervised by M. Akian, M. Bouhtou, and S. Gaubert, and supported by a research contract (CRE) of Orange Labs, was to apply optimization and game theory methods to the optimization of online services. We started by investigating the problem of the optimization of referencing, which we modelled by considering a family of compulsory hyperlinks and resources (fixed in advance by the website designer, these constitute the "skeletton" of the website) and also a family of facultative hyperlink or resources (some links may be ommited or some other links may be added).

In [20], we are approaching the problem of the pagerank optimization in this framework, by applying usual and constrained Markov decision processes techniques. This problem can indeed be reduced to an ergodic control problem without or with (ergodic) constraints, depending on the fact that hyperlinks constraints are local to each web page or depend on several web pages. We study the relaxed problem where the transition probabilities from one page to another may be any positive real (one may assume for instance that this probability depends on the position and type used for the corresponding hyperlink), as well as the discrete problem where these probabilities are uniform among the positive ones (as in the usual modelisation leading to the Pagerank). We show that these problems can be reduced to dynamic programming problems with exponentially many discrete actions, in which however the polytopes of transition probability measures admit polynomial time separation oracles. We derive from this approach polynomial time complexity results, as well as under some additional assumption, scalable algorithms (adapted to large web graphs), coupling dynamic programming and Lagrange relaxation. The latter have been tested on a real subgraph of the web.

A classical alternative ranking relies on the Perron eigenvector, as in the case of the algorithm "HITS" by Kleinberg. O. Fercoq treated the associated optimisation problem, which turns out to be harder than in the pagerank case, due to the lack of convexity properties. However, he developed a fast (sparse) algorithm, exploiting the rank 1 properties of operators appearing when computing the derivative of the objective function, allowing one to compute a local optimum [106].

O. Fercoq also developed a similar method to optimize Tomlin's "HOTS" score [38]. However, the convergence of the original HOTS algorithm was not proved. In [50], O. Fercoq has computed the convergence rate of the algorithm and of several of its variants, using techniques of nonexpansive mappings and properties of problems of flow with maximal entropy in a network.

The PhD thesis of Fercoq also comprises a ranking algorithm allowing one to detect spam pages [33] among dubious pages, starting from a seed (set of pages which are surely known to be spam). This algorithm relies on the Pagerank optimization techniques of [33].

6.5.6. Gestion du revenu appliquée à la tarification de services données/Yield management applied to pricing of data services

Participants: Mustapha Bouhtou [Orange Labs], Jean-Baptiste Dumont, Stéphane Gaubert.

Le travail de thèse CIFRE de J-B. Dumont, supervisée par M. Bouhtou et S. Gaubert, porte sur la tarification de services data et la gestion des ressources dans les réseaux mobiles. Celle-ci est abordée à l'aide de techniques de contrôle et d'optimisation stochastique. Dumont a développé un modèle de tarification, permettant d'analyser des mécanismes incitant les clients à reporter leur demande en dehors des periodes les plus chargées.

English version

The CIFRE PhD work of J-B. Dumont is jointly supervised by M. Bouhtou and S. Gaubert. It deals with the pricing of data services and resource allocation in mobile networks. This is addressed through stochastic control and stochastic optimization techniques. Dumont developed a model of pricing, in order to analyse incitations for customers to move their demand from loaded to less loaded time periods.

6.5.7. Vérification de systèmes temps-réels/Verification of real-time systems

Participants: Xavier Allamigeon, Uli Fahrenberg [IRISA], Stéphane Gaubert, Ricardo Katz [Conicet], Axel Legay [IRISA], Søren Ravn [Aalborg University].

Dans [140], Lu, Madsen, Milata, Ravn, Fahrenberg et Larsen ont montré que les polyèdres tropicaux peuvent être utilisés dans le cadre de l'analyse d'accessibilité d'automates temporisés. En effet, les polyèdres tropicaux expriment naturellement des invariants non-convexes, qui sont en fait des disjonctions d'invariants fournis par des DBM (*difference bound matrices*). A ce titre, les polyèdres tropicaux devraient permettre de réduire le nombre de disjonctions réalisées pendant l'analyse d'automates temporisés. Une limitation importante de cette approche est cependant que les polyèdres tropicaux sont topologiquement fermés, et qu'ils ne peuvent donc pas exprimer de contraintes d'inégalités strictes. Ces dernières sont néanmoins fondamentales dans l'analyse de systèmes temps-réels.

Nous avons donc développé une généralisation des polyèdres tropicaux permettant d'exprimer des contraintes mixtes, *i.e.* strictes ou larges. Notre approche repose sur l'utilisation d'inégalités tropicales linéaires à coefficients dans un (quotient du) semi-anneau de germes affines. Afin de réaliser des opérations sur cette nouvelle classe de polyèdres tropicaux, nous avons défini deux nouveaux algorithmes. Le premier est un analogue tropical de l'élimination de Fourier-Motzkin. Celle-ci s'applique plus généralement à des systèmes d'inégalités linéaires sur des semi-anneaux idempotents et totalement ordonnés. Le second algorithme permet de tester si un système de contraintes mixtes admet une solution. Nous montrons en effet que ce problème est équivalent en temps polynomial à la résolution d'un problème de jeux déterministes à somme nulle. Ces deux contributions nous permettent de définir les primitives requises pour l'analyse d'accessibilité d'automates temporisés.

Un autre problème important survenant dans cette application est l'élimination rapide de vecteurs linéairement dépendants (au sens tropical). Pendant son stage à Inria et au CEA (avril-juillet 2012) supervisé par X. Allamigeon, S. Gaubert et E. Goubault, S. Ravn a implémenté un algorithme dont la complexité est reliée à la taille du résultat (*output-sentive complexity*). Il a également implémenté une interface entre la bibliothèque TPLib et l'outil VerifyTAPN (https://launchpad.net/verifytapn).

English version

Lu, Madsen, Milata, Ravn, Fahrenberg and Larsen have shown in [140] that tropical polyhedra can be applied to the reachability analysis of timed automata. Indeed, tropical polyhedra naturally express non-convex invariants, which correspond to disjunctions of invariants provided by DBM (*difference bound matrices*). Consequently, tropical polyhedra should allow to reduce the number of disjunctions arising during the analysis of timed automata. An important limitation of this approach is that tropical polyhedra are topologically closed, and thus they cannot express strict inequality constraints. However, such constraints plays an important role in the analysis of real-time systems.

As a result, we have developed a generalization of tropical polyhedra, in order to express mixed constraints, *i.e.* strict or loose ones. Our approach relies on tropical linear inequalities with coefficients in a (quotient of) the semiring of affine germs. In order to perform operations on this new class of polyhedra, we have introduced two new algorithms. The first one is a tropical analog of Fourier-Moztkin elimination. In fact, it applies more generally to systems of linear inequalities over totally ordered and idempotent semirings. The second algorithm allows to test the feasability of a mixed constraint system. We indeed show that this problem is polynomial-time equivalent to solving mean payoff games. These two contributions allow to define the primitives required by the reachability analysis of timed automata.

Another important problem arising in this application is the fast elimination of linearly dependent vectors (in the tropical sense). During its internship at Inria and CEA (April-July 2012) supervised by X. Allamigeon, S. Gaubert and E. Goubault, S. Ravn has implemented an output-sensitive algorithm to eliminate such vectors. He has also implemented an interface between the library TPLib and the model-checker VerifyTAPN (https://launchpad.net/verifytapn).

MC2 Project-Team

6. New Results

6.1. Multi-fluid flows

- Microfluidics : Participants: Charles-Henri Bruneau, Johana Pinilla (PhD), Sandra Tancogne (MCF Reims).
  To handle oil recovery by chemical processes it is useful to better understand the behaviour of multifluids flows in a saturated soil. The porous medium is mimiced by a network of micro channels. The simulation of immiscible multi-fluids flows is then performed by means of the level-sets and the penalization methods to track the interfaces between the fluids and to get rid of the geometry difficulties. In addition the Cox law is added in the model to better move the interfaces during the simulations.
  Concerning visco-elastic fluids in micro-channel, one has often to compute solutions of system for which the viscosity in the stokes part is much smaller than that involved in the extra-stress. In his thesis, V. Huber has constructed a second order scheme solving Stokes equations for a bifluid flow with surface tension on a cartesian grid using a mixte finite volume-finite element approach.

6.2. Cancer modelling

We have improved our generic mathematical models describing tumor growth. These models were then specialized for several types of cancer (thyroidal lung nodules, brain tumors). The algorithm used to recover the parameters of these models from medical images has also been greatly improved and is now adapted to run on HPC architectures.

- Secondary tumors in the lung:
  The mathematical models describing the growth of secondary in the lungs have now settled and are well understood. The main focus of the year was to keep on using these models on patient data. New clinical case were selected by clinicians from the Institut Bergonié, there are currently under study. The model is currently able to reproduce the growth observed on 5 clinical cases. In 2011, various improvements to the calibration algorithms were made. The initial seeding of the algorithms was a weak point of the procedure. This has been much improved using a genetic algorithm. A complete rewrite of the routines was done to improve their versatility and efficiency. Previously, the numerical simulations and calibration were performed in 2D (clinicians selected the most relevant slice showing the evolution of the tumor). Work is now ongoing to switch to full 3D computations and calibration.
- Metastasis to the liver of a GIST
  Gastro-Intestinal Stromal Tumors often create metastasis to the liver. We have modeled the response to the treatment of such lesion starting from CT-scans.
- Modeling glioblastomas:
  In 2011, a hierarchy of models describing the growth of brain tumors was developed (and described in a submitted paper) in collaboration with University of Alabama at Birmingham. As we wished to obtain models that could be calibrated from patient data and yet be reasonably accurate, we believe that these models are suitable trade-offs between the simplicity of the SwansonÕs model (the only one used on patient data of brain tumors so far) and the accuracy of more complex models (that cannot really produce quantitative results). In particular, two models were built. The first one allows to study the efficacy of anti-angiogenic therapies. It seems to predict that the efficacy of these treatments is limited, this could be confirmed by a world-wide ongoing clinical study. The second model has been validated and we are trying to recover its parameters for a patient in 3D (which is a rather unique initiative to our knowledge).
- Modelling of electrochemotherapy :

Two articles related to the electrical cell modelling have been done ([64], [61]) . The first one deals with the influence of the ionic fluxes on the transmembrane voltage potential and on the cell volume. The main insight of the results consists in linking the transmembrane potential with the cell volume: it has been observed experimentally that cells with a low voltage potential do divide, whereas cells with high voltage potential do not, and the obtained relationship between voltage potential and cell volume can provide an explanation. The second article deals with a new model of cell electroporation essentially based on the experimental results of the I.G.R. In this paper we describe precisely the model, which takes into account the main experimental results in the electroporation process, and we present a variationnal formulation inherent to the model that leads to new efficient schemes in order to numerically solve the involved P.D.E.
The article describing a new electrical model of classical has been accepted in Journal of Math Biology [27]. This new phenomenological model involves much less parameters than the usual models, but it still provides the qualitatively good description of the electroporation. The main feature of this model lies in the fact that it provides an intrinsic behavior of the cell membrane, which seems in accordance with the preliminary experimental results of the IGR partner. We also adapted the finite difference method developed by L. Weynans and M. Cisternino for elliptic interface problems to the electropermeabilization model developed recently by C. Poignard with O. Kavian. The new method has been validated by convergence tests and comparison with other models. We have proven that in one dimension the numerical solution converges to the solution of the exact problem.

- Cell Migration modelling:
  The collaboration with IECB (University of Bordeaux) has continued with the postdocatoral position of Julie Joie. We have obtain a continuous model of cell density evolving on micropatterned polymers. The research report RR 7998 will be published in Math. Biosci. and Eng. A discrete model describing the single cells motility is being written.

  We also have started a collaboration with the University of Osaka (Japan), thanks to a PHC Sakura project, on the invadopodia. C. Poignard has been invited at Osaka in februray by Prof. Suzuki and T.Colin and C.Poignard have been at Osaka in september. A model describing the destruction of the extracellular matrix by the MMP enzyme, and then the cell migration has been obtained. R. Mahumet, a PhD student of Prof. Suzuki is developing a code to simulate the model.

6.3. Newtonian fluid flows simulations and their analysis

- Simulations of water distribution systems :Water losses may constitute a large amount of the distributed total water volume throughout water distribution systems. Here, a new model method is proposed that intends to minimize the total water volume distributed through leakage reduction. Our group has worked on the derivation of advection-reaction-diffusion type equations with an explicit relationship between the local pressure and the leakage rate. An original splitting technique to solve this type of hydraulic problem was then achieved. This technique allows pressure-dependent leakage to be taken into account, whereas in most models leakage is assumed to be uniform along a pipe. Finally, a constrained optimization problem was formulated for leakage reduction in WDS. The control variable had the mean of a local head loss and is considered in the Boundary Conditions to avoid dealing with discontinuities in the governing equations. The objective function to minimize was a regularization of the total water volume distributed. Specific operational constraints were added to ensure enough pressure at consumption points. The direct solution for this minimization problem was sought with a Gradient type method. The leakage reduction was proven to be significant in a case study. The percentage of leakage reduced from 24% to 10% in the linear relationship between pressure and leakage flow rate. With other leakage exponents, the same rate of reduction was achieved . The method was applied on a real network in the South-West of France. Controlling the pressure at two different strategic points permits a significant amount of the total distributed water to be saved (5%). This work was performed in collaboration with Cemagref Bordeaux . Future work will consist of applying a sensibility analysis of control location points to optimize the method.

- Incompressible flows : modeling and simulation of moving and deformable bodies. The incompressible Navier-Stokes equations are discretized in space onto a fixed cartesian mesh. The deformable bodies are taken into using a first order penalization method and/or second order immersed boundary method. The interface between the solid and the fluid is tracked using a level-set description so that it is possible to simulate several bodies freely evolving in the fluid. A turbulence model based on Samgorinsky model has been added to the numerical code. The numerical code written in the C langage is massively parallel. The large linear systems (over than 100 millions of dofs) are solved using the Petsc Library. As an illustration of the methods, fish-like locomotion is analyzed in terms of propulsion efficiency. Underwater maneuvering and school swimming are also explored. We were able to simulate the three-dimensional flow about a swimmer for realistic physical configurations. Another application is the turbulent 3D flow around complex wind turbine (see http://www.math.u-bordeaux1.fr/~mbergman and http://www.math.u-bordeaux1.fr/MAB/mc2/analysis.html for simulation movies). Wake flows generated by boat propellers are also modeled and simulated.
  We recently take in account a simplified elasticity model of the swimmer (elastic caudal tail of a fish). Some elastic parameters allows to increase the swimming efficiency around 20%-30%. Recent developments on multiphase flows have been performed. We are able to simulate water/air interactions with interface regularization. The interface with a boat is also taken into account. See http://www.math.u-bordeaux1.fr/~mbergman for simulations.
- Turbulence flow on an hemisphere : Participants: Charles-Henri Bruneau, Patrick Fischer (MCF Bordeaux 1), Yong Liang Xiong (PostDoc)
  ANR Cyclobulle lead by Hamid Kellay Soap hemi-bubble film experiments have shown some links between the formation of vortices when the hemi-bubble is heated at the equator and the formation of tornados in the earth atmosphere. Two-dimensional simulations using a stereographic map are used to compare to these experimental results and confirm the results when Coriolis force and heat source terms are added.
- Compressible flows: Immersed boundary methods. We are concerned with immersed boundary methods, i.e., integration schemes where the grid does not fit the geometry, and among this class of methods, more specifically with cartesian grid methods, where the forcing accounting for the presence of boundaries is performed at the discrete level. We have developed a simple globally second order scheme inspired by ghost cell approaches to solve compressible flows, inviscid as well as viscous. In the fluid domain, away from the boundary, we use a classical finite-volume method based on an approximate Riemann solver for the convective fluxes and a centered scheme for the diffusive term. At the cells located on the boundary, we solve an ad hoc Riemann problem taking into account the relevant boundary condition for the convective fluxes by an appropriate definition of the contact discontinuity speed. This method can easily be implemented in existing codes and is suitable for massive parallelization. It has been validated in two dimensions for Euler and Navier-Stokes equations, and in three dimensions for Euler equations. The order of convergence is two in L^2 norm for all variables, and between one and two in L^∞ depending on the variables. The 3D code has been parallelized with MPI. The case of a moving solid has been tested (flapping wing) and gives results for the drag and the lift in agreement with the references in the literature.
  The Oldroyd B constitutive model is used to study the role of the viscoelasticity of dilute polymer solutions in two-dimensional flows past a bluff body using numerical simulations. This investigation is motivated by the numerous experimental results obtained in quasi two dimensional systems such as soap film channels. The numerical modeling is novel for this case and therefore a comprehensive comparison is carried out to validate the present penalization method and artificial boundary conditions. In particular we focus on flow past a circular object for various values of the Reynolds number, Weissenberg number, and polymer viscosity ratio. Drag enhancement and drag reduction regimes are discussed in detail along with their flow features such as the pattern of vortex shedding, the variation of lift as well as changes in pressure, elongational rates, and polymer stress profiles. A comprehensive study of the flow behavior and energy balance are carefully carried out for high

Reynolds numbers. Flow instabilities in both numerical and experimental results are discussed for high Weissenberg numbers .

- Elliptic problems: We have developed a new cartesian method to solve elliptic problems with immersed interfaces. These problems appear in numerous applications, among them: heat transfer, electrostatics, fluid dynamics, but also tumour growth modelling, or modelling of electric potential in biological cells This method is second order accurate in the whole domain, notably near the interface. The originality of the method lies on the use of additionnal unknows located on interface points, on which are expressed flux equalities. Special care is dedicated to the discretization near the interface, in order to recover a stable second order accuracy. Actually, a naive discretization could lead to a first order scheme, notably if enough accuracy in the discretization of flux transmission condtions is not provided. Interfaces are represented with a distance level-set function discretized on the grid points. The method has been validated on several test-cases with complex interfaces in 2D. A parallel version has been developed using the PETSC library.
- Simulations of fluid-solid interactions : The interaction of an elastic structure and an fluid occurs in many phenomena in physics. To avoid the difficulty of coupling lagrangian elasticity with an eulerian fluid we consider a whole eulerian formulation. The elasticity of the structure is computed with retrograde caracteristics which satisfy a vectorial transport equation. We derive the associated fluid-structure models for incompressible and compressible media. The equations are discretized on a cartesian mesh with finite differences and finite volumes schemes. The applications concern the bio-locomotions and the study of air-elastic interaction.
- Vortex methods : The aim of this work is to couple vortex methods with the penalization methods in order to take advantage from both of them. This immersed boundary approach maintains the efficiency of vortex methods for high Reynolds numbers focusing the computational task on the rotational zones and avoids their lack on the no-slip boundary conditions replacing the vortex sheet method by the penalization of obstacles. This method that is very appropriate for bluff-body flows is validated for the flow around a circular cylinder on a wide range of Reynolds numbers. Its validation is now extended to moving obstacles (axial turbine blades) and three-dimensional bluff-bodies (flow around a sphere). See [77]. Moreover, using the global properties of the penalization method, this technique permits to include porous media simultaneously in the flow computation. We aim to adapt the porous media flows to our new method and to apply it in order to implement passive control techniques using porous layers around bluff-bodies.
- Domain decomposition : Domain decomposition methods are a way to parallelize the computation of numerical solutions to PDE. To be efficient, domain decompositions methods should converge independently on the number of subdomains. The classical convergence result for the additive Schwarz preconditioner with coarse grid is based on a stable decomposition. The result holds for discrete versions of the Schwarz preconditioner, and states that the preconditioned operator has a uniformly bounded condition number that depends only on the number of colors of the domain decomposition, and the ratio between the average diameter of the subdomains and the overlap width. Constants are usually non explicit and are only asserted to depend on the "shape regularity" of the domain decomposition.
  two years ago, we showed the result holds the additive Schwarz preconditioner can also be defined at the continuous level and provided completely explicits estimates. Last year, we established that a similar result also holds for non shape regular domain decompositions where the diameter of the smallest subdomain is significantly smaller than the diameter of the largest subdomain. The constants are also given explicitely and are independent of the ratio between the diameter of the largest sudomain and the diameter of the smallest subdomain.
  This year, we have studied explored new coarse spaces algorithms for domain decomposition methods. Coarse spaces are necessary to get a scalable algorithm whose convergence speed does not deteriorate when the number of subdomains increases. For domains decomposition methods with discontinuous iterates, we showed that continuous coarse spaces can never be an optimal choice. As

an alternative, we introduced both the use of discontinuous coarse spaces(DCS) and a new coarse space algorithm using these discontinuous coarse spaces.

6.4. Flow control and shape optimization

- Flow control : Participants: Charles-Henri Bruneau, Iraj Mortazavi, Emmanuel Creusé (Lille), Patrick Gilliéron (Paris).

  An efficient active control of the two- and three-dimensions flow around the 25 degrees rear window Ahmed body has been performed. A careful theoretical and numerical study of the trajectories of the vortices allows to adapt the control in order to improve its efficiency and get a better drag reduction.

6.5. Calculation of Ice Chunk Trajectory

- Participants: Héloise Beaugendre, Ramesh Yapalparvi.
  In this work, calculation of trajectories of ice chunk are carried out at varying values of ratio of density of ice piece to that of the ambient fluid. Proper Orthogonal Decomposition with Interpolation (PODI) method is then applied on snapshots of trajectories simulated by computational fluid dynamics. Snapshots of trajectories are obtained based on cartesian grids, penalization, and level sets. The extracted POD modes from snapshots are then used to reconstruct solutions and capabilities of POD with interpolation are demonstrated on ice trajectory calculations for flow around iced airfoil and cylinder for density ratio's that are not part of the snapshot set.

MCTAO Team

5. New Results

5.1. Optimal control for quantum systems: the contrast problem in NMR

These studies aim at optimizing the contrast in Nuclear Magnetic Resonance imaging using advanced optimal control.

5.1.1. Theoretical aspects

Participants: Bernard Bonnard, John Marriott, Monique Chyba [University of Hawaii], Gautier Picot [University of Hawaii], Olivier Cots, Jean-Baptiste Caillau.

This is done in collaboration with University of Hawaii, and deals with many theoretical aspects of the contrast problem in NMR: analysis of the optimal flow, feedback classification in relation with the relaxation times of the species. This activity has been the object of two publications [5], [4], and a conference talk [14] on feedback classification in the contrast problem, that will be followed by a journal article.

John Marriott will defend his Phd thesis on this topic, august 28, 2013; This will be followed by a two day conference on quantum control systems with applications, supported by a NSF grant and by the Engineering Department (P.E. Crouch).

5.1.2. Experimental aspects

Participants: Bernard Bonnard, Olivier Cots, Dominique Sugny [Univ. de Bourgogne], Steffan Glaser [TU München].

As said in section 4.2 , our work on this problem is based on experiments conducted in Prof. S. Glaser in Munich. Experiments using our techniques and measuring the improvement between materials that have an importance in medicine, like oxygenated and de-oxygenated blood have been conducted successfully, see [7], [9].

5.1.3. Numerical aspects

Participants: Bernard Bonnard, Olivier Cots, Jean-Baptiste Caillau.

In december, Pierre Martinon and Mathieu Caeys visited our group. This launhes a collaboration whose objective is to compare the direct and indirect methods in the contrast problem (implemented in the Bocop and Hampath sofwares) and use LMI techniques to get a global bound on the problem (in the contrast problem there are many local optima and the global optimality is a complicated issue)-also O. Cots visited R. Zidani (COMMANDS team) to investigate the use of numerical HJB techniques in the problem. This collaboration will allow to compare in a physical important problem the various available numerical methods in optimal control.

5.2. Conjugate and cut loci computations and applications

Participants: Bernard Bonnard, Olivier Cots, Jean-Baptiste Caillau.

One of the most important results obtained by B. Bonnard and his collaborators concern the explicit computations of conjugate and cut loci on surfaces. This has applications in optimal control to compute the global optimum and in optimal transport where regularity properties of the map in the Monge problem is related to convexity properties of the tangent injectivity domains. This shows also the transverse part of the team: [3] complete the previous results obtained with Rifford [33] ; the paper [20] analyses the conjugate and cut loci in Serret-Andoyer metrics and dynamics of spin particles with Ising coupling, and is a first step towards the computation of conjugate and cut loci on left invariant Riemannian and SR- metrics in S0(3) with applications for instance to the attitude control problem of a spacecraft. The submitted paper [19] concerns the analysis of singular metrics on surfaces in relation with the average orbital transfer problem.

5.3. Averaging in control

Participants: Bernard Bonnard, Helen-Clare Henninger, Jean-Baptiste Pomet.

A reference paper on the construction and properties of an “average control system” [2] is to be published; it is based on Alex Bombrun’s doctoral work (2007). It connects solutions of highly oscillating control systems to those of an average control system, when the frequency of oscillation goes high.

This average system in the case of minimum time for low thrust orbit transfer in the two body problem is currently being explored, in particular the study of its inherent singularities. Helen Henninger’s PhD aims at going much further in this direction and then apply this local study to real missions, possibly in a three-body environment.

5.4. Optimal transport

Participants: Ludovic Rifford, Alice Erlinger, Ahed Hindawi, Alessio Figalli, Bernard Bonnard, Jean-Baptiste Caillau, Lionel Jassionesse, Robert Mc Cann [U. of Toronto].

This year has seen new results or starting directions in many areas of optimal cotrol.

- The very general condition for continuity of the transport map given in [47] motivated exploration of conditions for convexity of the tangent injectivity domain [10], [3] on. Lionel Jassionnesse’s PhD is in part devoted to Ma-Tudinger-Wang tensor that also plays an important role in this matter.
- In Ahed Hindawi’s PhD [1], defended this year, results in optimal transport for sub-Riemannian costs (see the survey [16]) are generalized to costs coming from optimal control problems with quadratic cost and a drift.
- Alice Erlinger’s PhD, joint with University of Toronto is exploring Optimal Transport’s application to modeling in economics

5.5. Applications of control methods to lDynamical systems

Participants: Ludovic Rifford, Ayadi Lazrag, Riccardo Ruggiero, Alessio Figalli, Rafael Ruggiero [PUC, Rio de Janeiro].

Ludovic Rifford and collaborators have been applying, with success, techniques from geometric control theory to open problems in dynamical systems. Mostly on genericity properties and using controllability methods to build suitable perturbations See [11], [13], [21].

Ayadi Lazrag’s PhD also deals with such problems

MICMAC Project-Team

5. New Results

5.1. Electronic structure calculations

Participants: Eric Cancès, Ismaila Dabo, Virginie Ehrlacher, David Gontier, Salma Lahbabi, Claude Le Bris, Gabriel Stoltz.

In electronic structure calculation as in most of our scientific endeavours, we pursue a twofold goal: placing the models on a sound mathematical grounding, and improving the numerical approaches.

E. Cancès, V. Ehrlacher, S. Lahbabi and G. Stoltz have addressed issues related to the modeling and simulation of defects in periodic crystals.

Computing the energies of local defects in crystals is a major issue in quantum chemistry, materials science and nano-electronics. In collaboration with M. Lewin (CNRS, Cergy), E. Cancès and A. Deleurence have proposed in 2008 a new model for describing the electronic structure of a crystal in the presence of a local defect. This model is based on formal analogies between the Fermi sea of a perturbed crystal and the Dirac sea in Quantum Electrodynamics (QED) in the presence of an external electrostatic field. The justification of this model is obtained using a thermodynamic limit of Kohn-Sham type models. In [24], E. Cancès and G. Stoltz have studied the time evolution of defects within this model, in the context of linear response, which allowed them to give a rigorous meaning to the Adler-Wiser formula for the frequency-dependent dielectric permittivity of crystals. In collaboration with M. Lewin, E. Cancès and S. Lahbabi have introduced in [54] a functional setting for mean-field electronic structure models of Hartree-Fock or Kohn-Sham types for disordered quantum systems, and used these tools to study the reduced Hartree-Fock model for a disordered crystal where the nuclei are classical particles whose positions and charges are random.

On the numerical side, E. Cancès has worked with Y. Maday and R. Chakir (University Paris 6) on the numerical analysis of the electronic structure models. In [22], they have obtained optimal *a priori* error bounds for the the planewave approximation of the Thomas-Fermi-von Weizsäcker and the Kohn-Sham LDA models. Together with Y. Maday, E. Cancès and V. Ehrlacher have analyzed the computation of eigenvalues in spectral gaps of locally perturbed periodic Schrödinger operators [23]. In [53], they have introduced a general theoretical framework to analyze non-consistent approximations of the discrete eigenmodes of a self-adjoint operator, focusing in particular on the discrete eigenvalues laying in spectral gaps. Applying this analysis to the supercell method for perturbed periodic Schrödinger operators, they derive optimal convergence rates for the planewave discretization method, taking numerical integration errors into account. These results, along with earlier work on greedy algorithms for nonlinear convex problems and the study of local defects in the Thomas-Fermi-von Weiszacker theory, are collected in [7].

In the work [38], Claude Le Bris, in collaboration with Pierre Rouchon (Ecole des Mines de Paris), has introduced a new efficient numerical approach, based on a model reduction technique, to simulate high dimensional Lindblad type equations at play in the modelling of open quantum systems. The specific case under consideration is that of oscillation revivals of a set of atoms interacting resonantly with a slightly damped coherent quantized field of photons. The approach may be employed for other similar equations. Current work is directed towards other numerical challenges for this type of problems.

5.2. Computational Statistical Physics

Participants: Matthew Dobson, Claude Le Bris, Frédéric Legoll, Tony Lelièvre, Francis Nier, Grigorios Pavliotis, Mathias Rousset, Gabriel Stoltz.

The extremely broad field of molecular dynamics is a domain where the MICMAC project-team, originally more involved in the quantum chemistry side, has invested a lot of efforts in the recent years. Molecular dynamics may also be termed computational statistical physics since the main aim is to numerically estimate average properties of materials as given by the laws of statistical physics. The project-team studies both deterministic and probabilistic techniques used in the field. One of the main difficulty is related to the metastable features of the generated trajectories: the system remains trapped over very long times in metastable states, which means that very long trajectories need to be generated in order to obtain macroscopically relevant quantities. This is related to the fact that the timescale at the microscopic level is much smaller than the timescale at the macroscopic level. In [66], we propose a summary of the mathematical approaches to quantify metastability, and which appear to be useful to analyze the numerical methods used in molecular dynamics.

5.2.1. Free Energy calculations

For large molecular systems, the information of the whole configuration space may be summarized in a few coordinates of interest, called reaction coordinates. An important problem in chemistry or biology is to compute the effective energy felt by those reaction coordinates, called free energy.

In the article [42], Tony Lelièvre, Mathias Rousset and Gabriel Stoltz study the application of constrained Langevin dynamics to the computation of free energy differences, by thermodynamic integration techniques and fluctuation relation (à la Jarzynski).

One interest of free energy computation techniques is that they appear to be useful in other fields, like in computational statistics where multimodal measures are also frequently encountered, so that standard Markov Chain Monte Carlo appraoches also suffer from metastability.

For example, in [25], Nicolas Chopin (CREST, ENSAE), T. Lelièvre and G. Stoltz explore the application of the Adaptive Biasing Force method to Bayesian inference. This sampling method belongs to the general class of adaptive importance sampling strategies which use the free energy along a chosen reaction coordinate as a bias. Such algorithms are very helpful to enhance the sampling properties of Markov Chain Monte Carlo algorithms, when the dynamic is metastable.

In [58], G. Fort (Telecom Paris), B. Jourdain (CERMICS), E. Kuhn (INRA), T. Lelièvre and G. Stoltz have considered the Wang-Landau algorithm. The authorshave proved that the Wang-Landau algorithm converges with an associated central limit theorem, and have provided an analysis of the efficiency of the algorithm in a metastable situation.

5.2.2. Convergence to equilibrium

An important question for the analysis of sampling techniques is the rate of convergence to equilibrium for stochastic trajectories.

In [65], F. Nier, T. Lelièvre and G. Pavliotis study the interest of using non-reversible stochastic dynamics to enhance the rate of convergence to equilibrium, compared to reversible dynamics. A well posed optimization problem is obtained and solved in the case of a linear drift for the overdamped Langevin dynamics.

5.2.3. Metropolis Hastings algorithms

A classical sampling tool used in molecular dynamics and in computational statistics is the Metropolis-Hastings algorithm. There has been a lot of work (see G. Roberts et al.) to study how the variance of the proposal should scale with the dimension of the problem, in order to optimize the sampling procedure. Most of these works assume that (i) the target probability is the product of n one dimensional laws and that (ii) the Markov chain starts at equilibrium.

In the two works [60], [59], T. Lelièvre and his co-authors have generalized these results when the initial distribution is not the target probability. The diffusive limit in the latte case is solution to a stochastic differential equation nonlinear in the sense of McKean. They have discussed practical counterparts in order to optimize the variance of the proposal distribution to accelerate convergence to equilibrium. The analysis confirms the interest of the constant acceptance rate strategy (with acceptance rate between $1/4$ and $1/3$) first suggested in the works of G. Roberts et al., at least for the Random Walk Metropolis algorithm.

5.2.4. Thermodynamic limit

The quasicontinuum method is an approach to couple an atomistic model with a coarse-grained approximation in order to compute the states of a crystalline lattice at a reduced computational cost compared to a full atomistic simulation.

In that framework, the team has addressed questions related to the *finite temperature* modeling of atomistic systems and derivation of coarse-grained descriptions, such as canonical averages of observables depending only on a few variables. In the one-dimensional setting, an efficient strategy that bypasses the simulation of the whole system had been proposed in 2010. We refer to [47] for a recent review. In collaboration with X. Blanc (Université Pierre et Marie Curie), F. Legoll has extended this strategy to the so-called membrane setting in [16].

When the temperature is small, a perturbation approach can be used to compute the canonical averages of these observables depending only on a few variables, at first order with respect to temperature. In collaboration with E. Tadmor, W. K. Kim, L. Dupuy and R. Miller, F. Legoll has analyzed such an approach in [46]. The numerical tests reported there show the efficiency of the approach, as long as the temperature is indeed small.

5.2.5. Sampling trajectories

There exist a lot of methods to sample efficiently Boltzmann-Gibbs distributions. The situation is much more intricated as far as the sampling of trajectories (and especially metastable trajectories) is concerned.

Following a numerical observation in a previous work on the sampling of reactive trajectories by a multilevel splitting algorithm, F. Cérou (Inria Rennes), A. Guyader (Inria Rennes), T. Lelièvre and F. Malrieu (Université de Rennes) study theoretically in [56] the distribution of the lengths of these trajectories, using large deviation techniques.

In [37], C. Le Bris and T. Lelièvre together with M. Luskin and D. Perez from Los Alamos National Laboratoy provide a mathematical analysis of the parallel replica algorithm, which has been proposed by A. Voter in 1997 to simulate very efficiently metastable trajectories. This work opens a lot of perspectives, by using a generic tool (the quasi stationary distribution) to make a link between a continuous state space dynamics (Langevin dynamics) and a discrete state space dynamics (kinetic Monte Carlo models).

In a work in progress, T. Lelièvre and F. Nier have studied the quasi-stationnary distribution in relation for an overdamped Langevin process in a bounded domain. In the small temperature limit and by making the connection with boundary Witten Laplacians, they are able to compute accurately the spatial exit law along the boundary and non perturbative accurate formulas when the potential is changed inside the domain.

5.2.6. Effective dynamics

For a given molecular system, and a given reaction coordinate $\xi : \mathbb{R}^n \mapsto \mathbb{R}$, the free energy completely describes the statistics of $\xi(X)$ when $X \in \mathbb{R}^n$ is distributed according to the Gibbs measure. On the other hand, obtaining a correct description of the dynamics along ξ is complicated.

F. Legoll and T. Lelièvre have introduced and analyzed some years ago a strategy to define a coarse-grained dynamics that approximates $\xi(X_t)$, when the state of the system X_t evolves according to the overdamped Langevin equation (which is ergodic for the Gibbs measure). We refer to [47] for a recent review. The aim was to get a coarse-grained description giving access to some *dynamical* quantities (and not only *equilibrium* quantities). Together with G. Samaey (KU Leuven), they have recently studied how to use this coarse-grained description, accurate when the time scale separation is asymptotically large, to somewhat precondition the dynamics of the actual system in cases when the time scale separation is not large. For that purpose, they have used the parareal framework, to iteratively correct the sequential coarse-grained trajectory by fine scale trajectories performed in parallel. The main difficulty is that the two models (the reference one and the coarse-grained one) do not act on the same variable: the reference model evolves all the variables, whereas the coarse-grained model only evolves the slow variables. As shown in [63] in a simplified context (that of singularly perturbed ODEs), the precise coupling between both models should be done carefully.

The above study is concerned with models with continuous state spaces. S. Lahbabi and F. Legoll have studied in [61] a related question in the framework of kinetic Monte Carlo models, where the state space is discrete. For some models involving some slow and some fast variables, the effective dynamics of the slow component has been identified, and a complete proof of convergence proposed.

5.2.7. Hamiltonian dynamics

Constant energy averages are often computed as long time limits of time averages along a typical trajectory of the Hamiltonian dynamics. One difficulty of such a computation is the presence of several time scales in the dynamics: the frequencies of some motions are very high (e.g. for the atomistic bond vibrations), while those of other motions are much smaller. This problem has been addressed in a two-fold manner.

Fast phenomena are often only relevant through their mean effect on the slow phenomena, and their precise description is not needed. Consequently, there is a need for time integration algorithms that take into account these fast phenomena only in an averaged way, and for which the time step is not restricted by the highest frequencies. In [29], M. Dobson, C. Le Bris, and F. Legoll have developed integrators for Hamiltonian systems with high frequencies. The integrators were derived using homogenization techniques applied to the Hamilton-Jacobi PDE associated to the Hamiltonian ODE. This work extends previous works of the team. The proposed algorithms can now handle the case when the (unique) fast frequency depends on the slow degrees of freedom, or when there are several fast constant frequencies.

Another track to simulate the system for longer times is to resort to parallel computations. An algorithm in that vein is the parareal in time algorithm. It is based on a decomposition of the time interval into subintervals, and on a predictor-corrector strategy, where the propagations over each subinterval for the corrector stage are concurrently performed on the processors. Using a symmetrization procedure and/or a (possibly also symmetric) projection step, C. Le Bris and F. Legoll, in collaboration with X. Dai and Y. Maday, have introduced several variants of the original plain parareal in time algorithm [28]. These variants, compatible with the geometric structure of the exact dynamics, are better adapted to the Hamiltonian context.

5.2.8. Nonequilibrium systems

The efficient simulation of molecular systems is known to be a much more complicated problem when the system is subjected to a non-conservative external forcing than when the system experiences conservative forces. Together with the sampling of metastable dynamics mentioned above, these are the two major research focus in molecular dynamics of the project-team.

Nonequilibrium molecular dynamics simulations can be used to compute the constitutive relation between the strain rate and stress tensor in complex fluids. This is fulfilled simulating molecular systems subject to a steady, non-zero macroscopic flow at a given temperature. Starting from a bath model, M. Dobson, F. Legoll, T. Lelièvre, and G. Stoltz have derived a Langevin-type dynamics for a heavy particle in a non-zero background flow [57]. The resulting dynamics, which is theoretically obtained when a *unique* large particle is considered, is numerically observed to also perform well when a *system* of many interacting particles within shear flow is considered.

Let us also mention that the article on the computation of the viscosity of fluids using steady state nonequilibrium dynamics with an external nongradient bulk forcing, in the framework of the PhD of Rémi Joubaud, has also been published [34]. In addition, the study by G. Stoltz and C. Bernardin on thermal transport in one-dimensional chains of oscillators whose kinetic and potential energy functions are the same, has been accepted and is now published [13].

5.3. Complex fluids

Participants: David Benoit, Sébastien Boyaval, Claude Le Bris, Tony Lelièvre.

In [41], Claude Le Bris and Tony Lelièvre review the state-of-the-art of numerical and mathematical results on micro-macro models for viscoelastic fluids.

Following previous works, in [32], Claude Le Bris and Tony Lelièvre together with Lingbing He analyze the longtime behaviour of nematic polymeric fluids (liquid crystals). The longtime asymptotic for such models is much richer than for flexible polymers, that were considered in a previous analysis. Indeed, for these models, periodic in time behaviours are observed.

In his PhD under the supervision of Claude Le Bris and Tony Lelièvre, David Benoît studies models of aging fluids developed at the ESPCI (Ecole supérieure de physique et de chimie industrielles) and designed to take into account phenomena such as shear thinning, aging and shear banding in falling sphere experiments. The work consists in studying on the one hand the mathematical well-posedness of some macroscopic models, see [51] and, on the other hand, in trying to understand the link between such macroscopic models and microscopic models which have been proposed to describe such fluids.

Related to the mathematical modelling of free-surface complex flows under gravity, a new reduced model for thin layers of a viscoelastic upper-convected Maxwell fluid was derived by S. Boyaval in collaboration with François Bouchut, and possibly discontinuous solutions were numerically simulated with a new finite-volume scheme of relaxation type that satisfies a discrete counterpart of the natural dissipation [20]. This work is being pursued for other models.

Finally, in [31], Alexandre Ern (CERMICS), Rémi Joubaud (CERMICS) and Tony Lelièvre analyze a model describing equilibrium binary electrolytes surrounded by charged solid walls. This work is done in collaboration with physicists from the group PECSA at Université Pierre et Marie Curie. Applications include the modelization of clays for the burying of nuclear waste.

5.4. Application of greedy algorithms

Participants: Sébastien Boyaval, Eric Cancès, Virginie Ehrlacher, Tony Lelièvre.

Greedy algorithms are used in many contexts for the approximation of high-dimensional functions: Proper Generalized Decomposition, Reduced Basis techniques, etc.

Various greedy algorithms for high-dimensional non-symmetric problems, and inherent theoretical and practical difficulties have been analyzed in [52]. Current research now aims at extending these techniques to the approximation of high-dimensional spectral problems. Prototypical applications include electronic structure calculations or the computation of buckling modes in mechanics.

In probabilistic methods for uncertainty quantification in mechanics, S. Boyaval has used a greedy algorithm to construct control variates for accelerating Monte-Carlo simuation in the cases where an expectation has to be computed many times [21]. The work is being applied to the uncertainty quantification in numerical models for hydraulic engineering.

Finally, in [55], Fabien Casenave (CERMICS), Alexandre Ern (CERMICS) and Tony Lelièvre study the influence of round-off errors on the evaluation of the a posteriori estimators in the reduced basis approach. In practice, the evaluation of the error estimator can become very sensitive to round-off errors. An explanation of this fact is proposed, as well as efficient remedies.

5.5. Mathematical Physics

Participant: Francis Nier.

In [10], A. Aftalion and F. Nier answer questions asked by J. Dalibard about the feasibility of artificial gauge potentials. This analysis provides the range of small parameters within which the linear adiabatic argument used by the physicists is certainly not destroyed by the non linear effects.

In [43], D. Le Peutrec, F. Nier and C. Viterbo give an accurate Arrhenius law for Witten Laplacian acting on p-forms. In the case of functions the exponentially small eigenvalues are given by exponentiated differences of enegy levels between local minima and saddle points (Arrhenius law). In the case of p-forms the association of critical points with index p and critical points with index p+1 or p-1, is more subtle and is provided by Barannikov's presentation of Morse theory.

In [11], Z. Ammari and F. Nier have proved the mean field dynamics of general bosonic systems in the presence of singular pair interaction potentials, including the important 3 dimensional Coulombic case. As compared with their previous works, they developed a slightly new strategy relying on measure transportation techniques and results presented by Ambrosio-Gigli-Savaré in their book "Gradient Flows: In Metric Spaces And In The Space Of Probability Measures" (2005).

5.6. Homogenization and related topics

Participants: Ronan Costaouec, Claude Le Bris, Frédéric Legoll, William Minvielle, Mathias Rousset, Florian Thomines.

The homogenization of (deterministic) non periodic systems is a well known topic. Although well explored theoretically by many authors, it has been less investigated from the standpoint of numerical approaches (except in the random setting). In collaboration with X. Blanc and P.-L. Lions, C. Le Bris has introduced in [17] a possible theory, giving rise to a numerical approach, for the simulation of multiscale nonperiodic systems. The theoretical considerations are based on earlier works by the same authors (derivation of an algebra of functions appropriate to formalize a theory of homogenization). The numerical endeavour is completely new. Promising results have been obtained on a simple case of a periodic system perturbed by a localized defect. Ongoing works consider other configurations, such as for instance an interface between two different crystalline phases.

A theme closely related to homogenization theory and on which several members of the project team have worked a lot in the past few years is the passage from discrete (atomistic) mechanics to continuum mechanics. In this direction, C. Le Bris, in collaboration with X. Blanc and P.-L. Lions, has established in [18] the rigorous continuum limit of the Newton equations of motion for some simple cases of one-dimensional atomistic system.

The project-team also has pursued its efforts in the field of stochastic homogenization of elliptic equations, aiming at designing numerical approaches that both are pratically relevant and keep the computational workload limited.

An interesting case in that context is when the randomness comes as a *small* perturbation of the deterministic case. As previously shown by earlier works of the project-team, this situation can indeed be handled with a dedicated approach, which turns out to be far more efficient than the standard approach of stochastic homogenization. A final component of the work completed by Florian Thomines during his PhD thesis has concerned the application of Reduced Basis techniques to that specific context of weakly stochastic homogenization problems. In particular, the approach has been adapted in [39] to efficiently compute the terms of the expansion previously developed by A. Anantharaman and C. Le Bris to approximate a certain category of weakly random homogenization problems. It has been demonstrated that the reduced basis technique is very helpful in this particular context and indeed allows for a speed up of the computation. Another application of the same technique, for a slightly different category of models (still in the framework of weakly random homogenization problems) originally derived by X. Blanc, P.-L. Lions and C. Le Bris, has also been explored. The difficulty, there, is to compute the various corrector equations that parametrically depend on the macroscopic location of the microstructure and the particular realization of that microstructure. The problem is definitely amenable to reduced basis techniques, as demonstrated by some preliminary tests, but definite conclusions on the general validity of the approach are yet to be obtained.

The team has also proceeded to address, from a numerical viewpoint, the case when the randomness is not small. In that case, using the standard homogenization theory, one knows that the homogenized tensor, which is a deterministic matrix, depends on the solution of a stochastic equation, the so-called corrector problem, which is posed on the *whole* space $\mathbb{R}^d$. This equation is therefore delicate and expensive to solve. In practice, the space $\mathbb{R}^d$ is truncated to some bounded domain, on which the corrector problem is numerically solved. In turn, this yields a converging approximation of the homogenized tensor, which happens to be a *random* matrix. For a given truncation of $\mathbb{R}^d$, the team has shown in [14] that the variance of this matrix can be reduced using the technique of antithetic variables. F. Legoll and W. Minvielle are currently extending this technique to nonlinear, convex homogenization problems.

In addition, C. Le Bris, F. Legoll, W. Minvielle and M. Rousset are currently investigating the possibility to use other variance reduction approaches, such as control variate techniques. A promising idea is to use the weakly stochastic model previously introduced by A. Anantharaman and C. Le Bris (in which a periodic model is perturbed by a *rare* stochastic perturbation) to build a control variate model. The preliminary results that have already been obtained are very encouraging.

Another contribution in stochastic homogenization is the following. C. Le Bris, in collaboration with X. Blanc and P.-L. Lions, has recently introduced a variant of the classical random homogenization. For that variant, as often in random homogenization, the homogenized matrix is again defined from a so-called corrector function, which is the solution to a problem set on the entire space. F. Legoll and F. Thomines have described and proved the almost sure convergence of an approximation strategy based on truncated versions of the corrector problem in [64]. F. Legoll and F. Thomines have also established, in the one-dimensional case, a convergence result on the residual process, defined as the difference between the solution to the highly oscillatory problem and the solution to the homogenized problem.

From a numerical perspective, the Multiscale Finite Element Method is a classical strategy to address the situation when the homogenized problem is not known (e.g. in difficult nonlinear cases), or when the scale of the heterogeneities, although small, is not considered to be zero (and hence the homogenized problem cannot be considered as an accurate enough approximation). The extension of this strategy to the stochastic case, when the tensor describing the properties of the material is the sum of a periodic term and a small random term, has been studied by C. Le Bris, F. Legoll and F. Thomines [36]. A method with a much smaller computational cost than the original MsFEM in the stochastic setting has been proposed. Provided the stochastic perturbation is indeed small, the proposed method is as accurate as the original one. The work [36] also provides a complete analysis of the approach, extending that available for the deterministic setting. Such analysis often rely on the rate of convergence of the two scale expansion (in the sense of homogenization theory) of the solution to the highly oscillatory elliptic partial differential equation. Such a result is classic for periodic homogenization. In generic stochastic homogenization, the rate can be arbitrary small, depending on the rate with which the correlations of the random coefficient vanish. C. Le Bris, F. Legoll and F. Thomines have established in [40] such a result for *weakly stochastic homogenization*, using asymptotic properties of the Green function of the elliptic operator $Lu = -\text{div}\ (A\nabla u)$ (where A is a periodic, coercive and bounded matrix), established by F. Legoll in collaboration with X. Blanc and A. Anantharaman [15].

Still in the framework of the Multiscale Finite Element approach, F. Thomines has further investigated, in collaboration with Y. Efendiev and J. Galvis (Texas A&M University), the use of Reduced Basis methods. They have considered an extension of the MsFEM approach, well suited to the high contrast case, i.e. the case when the ratio between the maximum and the minimum values of the heterogeneous coefficient is large. The main idea of this extension is to complement the standard MsFEM basis functions with the eigenfunctions (associated to the first small eigenvalues) of a local eigenvalue problem. In [30], Y. Efendiev, J. Galvis and F. Thomines have considered the case when the problem depends on an additional parameter, and have shown how to use the Reduced Basis approach to more efficiently compute the eigenfunctions mentioned above.

Even in simple deterministic cases, there is actually still room for improvement in many different directions for the MsFEM approach. In collaboration with A. Lozinski (University of Toulouse and now at the University of Besançon) who visited the team-project repeatedly during the year, F. Legoll and C. Le Bris have introduced and studied a variant of MsFEM that considers Crouzeix-Raviart type elements on each mesh element. The continuity across edges (or facets) of the (multiscale) finite element basis set functions is enforced only weakly, using fluxes rather than point values. The approach has been analyzed (combining classical arguments from homogenization theory and finite element theory) and tested on simple, but highly convincing cases [35]. In particular, an elliptic problem set on a domain with a huge number of perforations has been considered in [62]. The variant developed outperforms all existing variants of MsFEM. A follow up on this work, in collaboration with U. Hetmaniuk (University of Washington in Seattle, two-week visitor in the project-team in the Spring of 2012), consists in the study of multiscale advection-diffusion problems. Such problems are possibly advection dominated and a stabilization procedure is therefore required. How stabilization interferes with the multiscale character of the equation is an unsolved mathematical question worth considering for numerical purposes.

Still another question related to homogenization theory that is investigated in the group is the following. Consider an elliptic equation, say in divergence form, with a highly oscillatory matrix coefficient. Is it possible to approximate the boundary value problem for different right hand sides using a similar problem with a *constant* matrix coefficient? How can this “best” constant matrix approximating the oscillatory problem be constructed in an efficient manner? How is this matrix related to the homogenized matrix, in the limit of infinitely rapidly oscillatory coefficients? Current work is directed towards solving such issues.

5.7. Asymptotic variance reduction

Participant: Mathias Rousset.

Recently, M. Rousset has initiated a research topic on variance reduction techniques (called ”asymptotic”) for the simulation of stochastic models of particles. The point is to use a macroscopic (or model reduced) equation as a control variate; or in other words, to use the information of a macroscopic description to decrease the statistical error of the simulated microscopic evolution.

A first step in this program has been achieved for a microscopic model describing the individual motion of bacteriae with a Markovian velocity-jump process. The macroscopic equation is an advection-diffusion equation called the chemotaxis equation. In [44], the pobabilistic derivation of the chemotaxis equation from the individual motion of bacteriae have been carried out in a rigorous way. In [45], a numerical method simulating the individual evolution of bacteriae with ”asymptotic” variance reduction have been proposed.

5.8. Computational materials spectroscopy in electrochemistry and optoelectronics

Participant: Ismaila Dabo.

Many advances in the understanding and design of nanomaterials have been enabled by spectroscopic techniques of increasing spatial and temporal resolution. In electrochemistry and optoelectronics, spectroscopy provides insight into the chain of processes involved in harnessing, storing, and delivering energy.

In support to experimental techniques, much progress has been achieved in simulating spectroscopic phenomena to shed light into energy conversion at the molecular scale. Such understanding is critical to the molecular design of a range of electrical devices, including but not limited to fuel cells, batteries, dye-sensitized solar cells, and optoelectronic devices.

The work of I. Dabo is dedicated to the development of quantum and semiclassical methods to simulate spectroscopies of electrochemical and optoelectronic materials. The achieved level of efficiency and accuracy fosters dialogue between experiment and theory for interpreting complex spectroscopic data. This year, these novel methods have been applied to simulate spectroscopic phenomena spanning the infrared to the visible and ultraviolet ranges.

The first application pertains to the infrared sum-frequency-generation (SFG) spectroscopy of adsorption mechanisms at the origin of the tolerance of fuel-cell catalytic electrodes to chemical poisoning. The study explains the critical influence of the electrode voltage in analyzing surface spectroscopy experiments (work done in collaboration with EPFL). [12], [26], [19]

The second application aims at understanding the sensitizing properties of organometallic dyes in dye-sensitized solar cells by simulating optical photoluminescence (PL) spectra, thereby elucidating the role of electron localization and ligand functionalization on the phosphorescence of organometallic complexes (work done in collaboration with the University of Minnesota). [33]

The third application is focused on the ultraviolet photoelectron spectroscopy (UPS) of photoactive nanomaterials of relevance to the design of organic photovoltaic junctions and photoelectrodes (work done in collaboration with the Italian Institute of Nanoscience, Seoul National University, and Xiamen University). [27]

Future challenges and opportunities are related to the time-dependent simulation of transient and cyclic spectra. These developments, which will be part of the widely used Quantum-ESPRESSO distribution (http://www.quantum-espresso.org), would pave the way for comprehensive studies of kinetic processes in tandem with time-resolved spectroscopic experiments.

MISTIS Project-Team

6. New Results

6.1. Mixture models

6.1.1. Taking into account the curse of dimensionality

Participant: Stéphane Girard.

Joint work with: Bouveyron, C. (Université Paris 1), Fauvel, M. (ENSAT Toulouse)

In the PhD work of Charles Bouveyron (co-advised by Cordelia Schmid from the Inria LEAR team) [53], we propose new Gaussian models of high dimensional data for classification purposes. We assume that the data live in several groups located in subspaces of lower dimensions. Two different strategies arise:

- the introduction in the model of a dimension reduction constraint for each group
- the use of parsimonious models obtained by imposing to different groups to share the same values of some parameters

This modelling yields a new supervised classification method called High Dimensional Discriminant Analysis (HDDA) [4]. Some versions of this method have been tested on the supervised classification of objects in images. This approach has been adapted to the unsupervised classification framework, and the related method is named High Dimensional Data Clustering (HDDC) [3]. Also, the description of the R package is published in [11]. Our recent work consists in adding a kernel in the previous methods to deal with nonlinear data classification [27], [45].

6.1.2. Robust mixture modelling using skewed multivariate distributions with variable amounts of tailweight

Participants: Florence Forbes, Darren Wraith.

Clustering concerns the assignment of each of N, possibly multidimensional, observations $y_1, ..., y_N$ to one of K groups. A popular way to approach this task is via a parametric finite mixture model. While the vast majority of the work on such mixtures has been based on Gaussian mixture models in many applications the tails of normal distributions are shorter than appropriate or parameter estimations are affected by atypical observations (outliers). In such cases, the multivariate student t distribution is motivated as a heavy-tailed alternative to the multivariate Gaussian distribution. The additional flexibility of the multivariate t comes from introducing an additional degree of freedom parameter (dof) which can be viewed as a robust tuning parameter.

A useful representation of the t-distribution is as a so-called *infinite mixture of scaled Gaussians* or *Gaussian scale mixture*,

$$p(y;\mu,\Sigma,\theta) = \int_0^\infty \mathcal{N}_M(y;\mu,\Sigma/w)\; f_W(w;\theta)\;\mathrm{d}w \tag{9}$$

where $\mathcal{N}_M(\,.\,;\mu,\Sigma/w)$ denotes the M-dimensional Gaussian distribution with mean μ and covariance Σ/w and f_W is the probability distribution of a univariate positive variable W referred to as the weight variable. When f_W is a Gamma distribution $\mathcal{G}(\nu/2,\nu/2)$ where ν denotes the degrees of freedom, we recover the multivariate t distribution. The weight variable W in this case effectively acts to govern the tail behaviour of the distributional form from light tails ($\nu \to \infty$) to heavy tails ($\nu \to 0$) depending on the value of ν.

For many applications, the distribution of the data may also be highly asymmetric in addition to being heavy tailed (or affected by outliers). A natural extension to the Gaussian scale mixture case is to consider *location* ***and*** *scale Gaussian mixtures* of the form,

$$p(y;\mu,\Sigma,\theta)=\int_0^\infty \mathcal{N}_M(y;\mu+w\beta\Sigma,w\Sigma)\; f_W(w;\theta)\;\mathrm{d}w, \tag{10}$$

where β is an additional M-dimensional vector parameter for skewness and the determinant of Σ equals 1 for parameter identifiability. When f_W is a Generalized Inverse Gaussian distribution ($GIG(y;\lambda,\delta,\gamma)$, we recover the family of Generalized Hyperbolic (GH) distributions. Depending on the parameter choice for the GIG, special cases of the GH family, include: the multivariate GH distribution with hyperbolic margins ($\lambda=1$); the normal inverse Gaussian distribution ($\lambda=-1/2$); the multivariate hyperbolic ($\lambda=\frac{M+1}{2}$) distribution; the hyperboloid distribution ($\lambda=0$); the hyperbolic skew-t distribution ($\lambda=-\nu,\gamma=0$); and the normal gamma distribution ($\lambda>0,\mu=0,\delta=0$)) amongst others. For applied problems, the most popular of these forms appears to be the Normal Inverse Gaussian (NIG) distribution, with extensive use in financial applications. Another distributional form allowing for skewness and heavy or light tails includes different forms of the multivariate skew-t. Most of these distributional forms are also able to be represented as *location and scale Gaussian mixtures*.

Although the above approaches provide for great flexibility in modelling data of highly asymmetric and heavy tailed form the above approaches assume f_W to be a univariate distribution and hence each dimension is governed by the same amount of tailweight. There have been various approaches to address this issue in the statistics literature for both symmetric and asymmetric distributional forms. In his work, [66] proposes a dependent bivariate t-distribution with marginals of different degrees of freedom but the tractability of the extension to the multivariate case is unclear. Additional proposals are reviewed in chapters 4 and 5 of [67] but these formulations tend to be appreciably more complicated, often already in the expression of the probability density function. Increasingly, there has been much research on copula approaches to account for flexible distributional forms but the choice as to which one to use in this case and the applicability to (even) moderate dimensions is also not clear. In general the papers take various approaches whose relationships have been characterized in the bivariate case by [73]. However, most of the existing approaches suffer either from the non-existence of a closed-form pdf or from a difficult generalization to more than two dimensions.

In this work, we show that the location and scale mixture representation can be further explored and propose a framework that is considerably simpler than those previously proposed with distributions exhibiting interesting properties. Using the normal inverse Gaussian distribution (NIG) as an example, we extend the standard *location and scale mixture of Gaussian representation* to allow for the tail behaviour to be set or estimated differently in each dimension of the variable space. The key elements of the approach are the introduction of multidimensional weights and a decomposition of the matrix Σ in (6) which facilitates the separate estimation and also allows for arbitrary correlation between dimensions. We outline an approach for maximum likelihood estimation of the parameters via the EM algorithm and explore the performance of the approach on several simulated and real data sets in the context of clustering.

6.1.3. Robust clustering for high dimensional data

Participants: Florence Forbes, Darren Wraith, Minwoo Lee.

For a clustering problem, a parametric mixture model is one of the popular approaches. Most of all, Gaussian mixture models are widely used in various fields of study such as data mining, pattern recognition, machine learning, and statistical analysis. The modeling and computational flexibility of the Gaussian mixture model makes it possible to model a rich class of density, and provides a simple mathematical form of cluster models.

Despite the success of Gaussian mixtures, the parameter estimations can be severely affected by outliers. By adding an additional degrees of freedom (dof) parameter, a robustness tuning parameter, the robust improvement in clustering has been achieved. Although adopting t distribution loses the closed-form solution, it is still tractable by representing t distribution as Gaussian scale mixture (GSM), which consists of a Gaussian random vector that is weighted by a hidden scaling variable. Recent work that uses the multivariate t distribution has showed the improved robustness.

Along with robustness from t distribution, for the practical use, efficient handling of a high dimensional data is critical. High dimensional data often make most of clustering methods perform poorly. To overcome the curse of dimensionality, Bouveyron et al. [54] proposed the model-based high dimensional data clustering (HDDC). HDDC searches the intrinsic dimension of each class with the BIC criterion or the scree-test of Cattell; this allows them to limit the number of parameters by taking into account only the specific subspace that each class is located. The parameterization makes HDDC not only computationally efficient but robust with respect to the ill-conditioning or the singularity of empirical covariance matrix.

This work proposes an approach that combines robust clustering with the HDDC. The use of the mixture of multivariate t distribution on the basis of HDDC develops robust high dimensional clustering methods that can capture various kinds of density models. Further, extending the mixture model with multiple t distributions for each dimension, we propose more flexible model that can be applicable to various data. We suggest a model-based approach for this method.

6.1.4. Partially Supervised Mapping: A Unified Model for Regression and Dimensionality Reduction

Participant: Florence Forbes.

Joint work with: Antoine Deleforge and Radu Horaud from the Inria Perception team.

We cast dimensionality reduction and regression in a unified latent variable model. We propose a two-step strategy consisting of characterizing a non-linear *reversed* output-to-input regression with a generative piecewise-linear model, followed by Bayes inversion to obtain an output density given an input. We describe and analyze the most general case of this model, namely when only some components of the output variables are observed while the other components are latent. We provide two EM inference procedures and their initialization. Using simulated and real data, we show that the proposed method outperforms several existing ones.

6.1.5. Variational EM for Binaural Sound-Source Separation and Localization

Participant: Florence Forbes.

Joint work with: Antoine Deleforge and Radu Horaud from the Inria Perception team.

We addressed the problem of sound-source separation and localization in real-world conditions with two microphones. Both tasks are solved within a unified formulation using supervised mapping. While the parameters of the direct mapping are learned during a training stage that uses sources emitting white noise (calibration), the inverse mapping is estimated using a variational EM formulation. The proposed algorithm can deal with natural sound sources such as speech which are known to yield sparse spectrograms, and is able to locate multiple sources both in azimuth and in elevation. Extensive experiments with real data show that the method outperform state-of-the-art both in separation and localization.

6.2. Statistical models for Neuroscience

6.2.1. Variational approach for the joint estimation-detection of Brain activity from functional MRI data

Participants: Florence Forbes, Lotfi Chaari, Thomas Vincent.

Joint work with: Michel Dojat (Grenoble Institute of Neuroscience) and Philippe Ciuciu from Neurospin, CEA in Saclay.

In standard within-subject analyses of event-related fMRI data, two steps are usually performed separately: detection of brain activity and estimation of the hemodynamic response. Because these two steps are inherently linked, we adopt the so-called region-based Joint Detection-Estimation (JDE) framework that addresses this joint issue using a multivariate inference for detection and estimation. JDE is built by making use of a regional bilinear generative model of the BOLD response and constraining the parameter estimation by physiological priors using temporal and spatial information in a Markovian model. In contrast to previous works that use Markov Chain Monte Carlo (MCMC) techniques to sample the resulting intractable posterior distribution, we recast the JDE into a missing data framework and derive a Variational Expectation-Maximization (VEM) algorithm for its inference. A variational approximation is used to approximate the Markovian model in the unsupervised spatially adaptive JDE inference, which allows automatic fine-tuning of spatial regularization parameters. It provides a new algorithm that exhibits interesting properties terms of estimation error and computational cost compared to the previously used MCMC-based approach. Experiments on artificial and real data show that VEM-JDE is robust to model mis-specification and provides computational gain while maintaining good performance in terms of activation detection and hemodynamic shape recovery. Main corresponding paper [13]

6.2.2. Hemodynamic-informed parcellation of fMRI data in a Joint Detection Estimation framework

Participants: Florence Forbes, Lotfi Chaari, Thomas Vincent.

Joint work with: Philippe Ciuciu from Team Parietal and Neurospin, CEA in Saclay.

Identifying brain hemodynamics in event-related functional MRI (fMRI) data is a crucial issue to disentangle the vascular response from the neuronal activity in the BOLD signal. This question is usually addressed by estimating the so-called Hemodynamic Response Function (HRF). Voxelwise or region-/parcelwise inference schemes have been proposed to achieve this goal but so far all known contributions commit to pre-specified spatial supports for the hemodynamic territories by defining these supports either as individual voxels or a priori fixed brain parcels. In this paper, we introduce a Joint Parcellation-Detection-Estimation (JPDE) procedure that incorporates an adaptive parcel identification step based upon local hemodynamic properties. Efficient inference of both evoked activity, HRF shapes and *supports* is then achieved using variational approximations. Validation on synthetic and real fMRI data demonstrate the JPDE performance over standard detection estimation schemes and suggest it as a new brain exploration tool. Corresponding papers [29], [28].

6.2.3. Variational variable selection to assess experimental condition relevance in event-related fMRI

Participants: Florence Forbes, Christine Bakhous, Lotfi Chaari, Thomas Vincent, Farida Enikeeva.

Joint work with: Michel Dojat (Grenoble Institute of Neuroscience) and Philippe Ciuciu from Neurospin, CEA in Saclay.

Brain functional exploration investigates the nature of neural processing following cognitive or sensory stimulation. This goal is not fully accounted for in most functional Magnetic Resonance Imaging (fMRI) analysis which usually assumes that all delivered stimuli possibly generate a BOLD response everywhere in the brain although activation is likely to be induced by only some of them in specific brain regions. Generally, criteria are not available to select the relevant conditions or stimulus types (e.g. visual, auditory, etc.) prior to activation detection and the inclusion of irrelevant events may degrade the results, particularly when the Hemodynamic Response Function (HRF) is jointly estimated. To face this issue, we propose an efficient variational procedure that automatically selects the conditions according to the brain activity they elicit. It follows an improved activation detection and local HRF estimation that we illustrate on synthetic and real fMRI data. This approach is an alternative to our previous approach based on Monte-Carlo Markov Chain (MCMC) inference [25]. Corresponding paper [26].

6.2.4. Bayesian BOLD and perfusion source separation and deconvolution from functional ASL imaging

Participants: Florence Forbes, Thomas Vincent.

In the context of ARC AINSI project, joint work with: Philippe Ciuciu from Neurospin, CEA in Saclay.

In many neuroscience applications, the Arterial Spin Labeling (ASL) fMRI modality arises as a preferable choice to the standard BOLD modality due to its ability to provide a quantitative measure of the Cerebral Blood Flow (CBF). Such a quantification is central but generally performed without consideration of a specific modeling of the perfusion component in the signal often handled via standard GLM approaches using the BOLD canonical response function as regressor. In this work, we propose a novel Bayesian hierarchical model of the ASL signal which allows activation detection and both the extraction of a perfusion and a hemodynamic component. Validation on synthetic and real data sets from event-related ASL show the ability of our model to address the source separation and double deconvolution problems inherent to ASL data analysis.

6.2.5. Extraction of physiological components in functional ASL data

Participants: Florence Forbes, Thomas Vincent, Lotfi Chaari, Marc Guillotin.

In the context of ARC AINSI project, joint work with: Jan Warnking (Grenoble Institute of Neuroscience) and Philippe Ciuciu from Neurospin, CEA in Saclay.

The internship of Marc Guillotin has been supported by Le pole Cognition de Grenoble.

The goal of this work was to investigate Independent component analysis techniques to identify the part of the ASL signal due to physiological sources such as respiratory and cardiac components. Once identified those physiological components should be removed to produce an uncontaminated ASL signal. This preliminary work showed that the physiological effects were affecting all signal components and were therefore not easy to extract without removing some of the useful signal. More experiments should be made on real data from the GIN.

6.2.6. Comparison of processing workflows for ASL data analysis

Participant: Thomas Vincent.

In the context of ARC AINSI project, joint work with: Michel Dojat (Grenoble Institute of Neuroscience), Philippe Ciuciu from Neurospin, CEA in Saclay, Remi Dubujet, Elise Bannier, Isabelle Courouge, Christian Barillot, Camille Maudet from EPI Visages in Rennes.

We assessed and compared the performance of different ASL processing pipelines in order to promote one using specific indexes (Contrast to noise ratio, partial volume effect, et). We proposed to assess the impact of the pipelines based on the quality of the final corrected ASL images using a common set of subjects for all workflows. We leaned on the expertise of the Visages and GIN teams on ASL, and first started from existing attempts made in the teams. At the moment, there is a striking lack of such guidelines. The recent toolbox ASLtbx proposes a number of procedures that are based on very standard tools (e.g. SPM) and do not make use of more efficient approaches from more recent literature. Similarly, in the BIRN project, processing pipelines are mentioned but none are currently available.

6.3. Markov models

6.3.1. Spatial risk mapping for rare disease with hidden Markov fields and variational EM

Participants: Florence Forbes, Senan James Doyle.

Joint work with: Lamiae Azizi, David Abrial and Myriam Garrido from INRA Clermont-Ferrand-Theix.

Current risk mapping models for pooled data focus on the estimated risk for each geographical unit. A risk classification, *i.e.* grouping of geographical units with similar risk, is then necessary to easily draw interpretable maps, with clearly delimited zones in which protection measures can be applied. As an illustration, we focus on the Bovine Spongiform Encephalopathy (BSE) disease that threatened the bovine production in Europe and generated drastic cow culling. This example features typical animal disease risk analysis issues with very low risk values, small numbers of observed cases and population sizes that increase the difficulty of an automatic classification. We propose to handle this task in a spatial clustering framework using a non standard discrete hidden Markov model prior designed to favor a smooth risk variation. The model parameters are estimated using an EM algorithm and a mean field approximation for which we develop a new initialization strategy appropriate for spatial Poisson mixtures. Using both simulated and our BSE data, we show that our strategy performs well in dealing with low population sizes and accurately determines high risk regions, both in terms of localization and risk level estimation.

Main corresponding paper [14].

6.3.2. Spatial modelling of biodiversity from high-througput DNA sequence data

Participants: Florence Forbes, Angelika Studeny.

This is joint work with Eric Coissac and Pierre Taberlet from LECA (Laboratoire d'Ecologie Alpine) and Alain Viari from EPI Bamboo

Biodiversity has been acknowledged as a vital ressource for ecosystem health and stability, faced with an unprecedented global decline. In order to be effective, conservation actions need to be based on reliable and fast analysis. Recent advances in DNA sequencing methods now enable DNA-based identification of multiple species from only few, even potentially degraded environmental samples (metabarcoding.org, [74]). This offers a new way of biodiversity assessment and is of particular interest where large-scale individual-based diversity assessment is difficult, for example in tropical environments. Due to their comparatively low demand in cost and effort, these methods are characterized by their high throughput; they are expected to produce vast amounts of data as they gain in popularity over the coming years. The specific properties of these data (e.g. bias from sequencing errors, notion of species) and their high dimensionality provides new statistical and computational challenges for biodiversity assessment. This project aims at extending existing summary statistics to be used with data from metabarcoding surveys and, where this is not adequate, to develop new methodology. A special focus is on the spatial mapping of biodiversity and the co-occurrence of species. In a first instance, we investigate spatial clustering algorithms based on Markov random fields (software SpaCEM3, http://spacem3.gforge.inria.fr/) to identify regions of high species occurrence as well as structured additive regression models and their implementation to estimate cross-correlations between species occurrences in space [61], [72], [71]. At present, results have been derived in form of species occurrence maps, which take into account pairwise cross-correlation, and interaction graphs.

6.3.3. Statistical characterization of tree structures based on Markov tree models and multitype branching processes, with applications to tree growth modelling.

Participant: Jean-Baptiste Durand.

Joint work with: Pierre Fernique (Montpellier 2 University and CIRAD) and Yann Guédon (CIRAD), Inria Virtual Plants.

The quantity and quality of yields in fruit trees is closely related to processes of growth and branching, which determine ultimately the regularity of flowering and the position of flowers. Flowering and fruiting patterns are explained by statistical dependence between the nature of a parent shoot (*e.g.* flowering or not) and the quantity and natures of its children shoots – with potential effect of covariates. Thus, better characterization of patterns and dependencies is expected to lead to strategies to control the demographic properties of the shoots (through varietal selection or crop management policies), and thus to bring substantial improvements in the quantity and quality of yields.

Since the connections between shoots can be represented by mathematical trees, statistical models based on multitype branching processes and Markov trees appear as a natural tool to model the dependencies of interest. Formally, the properties of a vertex are summed up using the notion of vertex state. In such models, the numbers of children in each state given the parent state are modeled through discrete multivariate distributions. Model selection procedures are necessary to specify parsimonious distributions. We developed an approach based on probabilistic graphical models to identify and exploit properties of conditional independence between numbers of children in different states, so as to simplify the specification of their joint distribution. The graph building stage was based on exploring the space of possible chain graph models, which required defining a notion of neighbourhood of these graphs. A parametric distribution was associated with each graph. It was obtained by combining families of univariate and multivariate distributions or regression models. These were chosen by selection model procedures among different parametric families.

This work was carried out in the context of Pierre Fernique's first year of PhD (Montpellier 2 University and CIRAD). It was applied to model dependencies between short or long, vegetative or flowering shoots in apple trees. The results highlighted contrasted patterns related to the parent shoot state, with interpretation in terms of alternation of flowering (see paragraph 6.3.4). It was also applied to the analysis of the connections between cyclic growth and flowering of mango trees. This work will be continued during Pierre Fernique's PhD thesis, with extensions to other fruit tree species and other parametric discrete multivariate families of distributions, including covariates and mixed effects.

6.3.4. Statistical characterization of the alternation of flowering in fruit tree species

Participant: Jean-Baptiste Durand.

Joint work with: Jean Peyhardi and Yann Guédon (Mixed Research Unit DAP, Virtual Plants team), Baptiste Guitton, Yan Holtz and Evelyne Costes (DAP, AFEF team), Catherine Trottier (Montpellier University)

The aim of this work was to characterize genetic determinisms of the alternation of flowering in apple tree progenies. Data were collected at two scales: at whole tree scale (with annual time step) and a local scale (annual shoot or AS, which is the portions of stem that were grown during the same year). Two replications of each genotype were available.

Indices were proposed to characterize alternation at tree scale. The difficulty is related to early detection of alternating genotypes, in a context where alternation is often concealed by a substantial increase of the number of flowers over consecutive years. To separate correctly the increase of the number of flowers due to aging of young trees from alternation in flowering, our model relied on a parametric hypothesis for the trend (fixed slopes specific to genotype and random slopes specific to replications), which translated into mixed effect modelling. Then, different indices of alternation were computed on the residuals. Clusters of individuals with contrasted patterns of bearing habits were identified.

To model alternation of flowering at AS scale, a second-order Markov tree model was built. Its transition probabilities were modelled as generalized linear mixed models, to incorporate the effects of genotypes, year and memory of flowering for the Markovian part, with interactions between these components.

Asynchronism of flowering at AS scale was assessed using an entropy-based criterion. The entropy allowed for a characterisation of the roles of local alternation and asynchronism in regularity of flowering at tree scale.

Moreover, our models highlighted significant correlations between indices of alternation at AS and individual scales.

This work was extended by the Master 2 internship of Yan Holtz, supervised by Evelyne Costes and Jean-Baptiste Durand. New progenies were considered, and a methodology based on a lighter measurement protocol was developed and assessed. It consisted in assessing the accuracy of approximating the indices computed from measurements at tree scale by the same indices computed as AS scale. The approximations were shown sufficiently accurate to provide an operational strategy for apple tree selection.

As a perspective of this work, patterns in the production of children ASs (numbers of flowering and vegetative children) depending on the type of the parent AS must be analyzed using branching processes and different types of Markov trees, in the context of Pierre Fernique's PhD Thesis (see paragraph 6.3.3).

6.4. Semi and non-parametric methods

6.4.1. Post-Reflow Automated Optical Inspection of Lead Defects

Participants: Florence Forbes, Kai Qin, Huu Giao Nguyen, Darren Wraith, Ludovic Leau-mercier.

This is joint work with VI-Technology in the context of the IVP project.

Quality and throughput in printed circuit board (PCB) assembly lines constitute a continuous challenge, especially when placing smaller components on boards that are becoming increasingly dense. Automated optical inspection (AOI) technology allows PCB assembly lines to keep operating at a high throughput while visually inspecting production quality in term of paste deposits, mounted components and solder joints in an automatic and non-contact manner. In the AOI, high definition cameras precisely move in both X- and Y-direction to scan the device under test lit by special lighting techniques, e.g. light-emitting diode (LED) lighting. The captured images are then analyzed using specific inspection algorithms to identify defects. The AOI systems can be placed at several stages during the manufacturing process, such as bare board inspection, solder paste inspection, pre-reflow inspection and post-reflow inspection, which usually need some time to be programmed via offline learning of verified boards and expert expertise before online inspection starts. Vi TECHNOLOGY (VIT) offers a wide range of AOI solutions to increase productivity throughout electronics manufacturing lines while enhancing the quality of products. Post-reflow AOI is implemented after the reflow procedure in PCB assembly lines to enable inspection of the major post-reflow defects. This work focus on certain types of post-reflow defects occurring on leaded components, i.e. lifted lead, no solder, excess of solder, contamination on lead, insufficient solder, bad wedding and dry joint. We aim at developing efficient post-reflow lead defect detection approaches by synergizing image analysis, pattern recognition, machine learning, and statistics techniques to improve performance of VIT commercial post-reflow AOI solutions from two aspects: 1) Reducing both detection escape rate and false detection rate; 2) Minimizing programming efforts. The exact nature of the work is confidential.

6.4.2. An Improved CUDA-Based Implementation of Differential Evolution on GPU

Participants: Kai Qin, Florence Forbes.

Modern GPUs enable widely affordable personal computers to carry out massively parallel computation tasks. NVIDIA's CUDA technology provides a wieldy parallel computing platform. Many state-of-the-art algorithms arising from different fields have been redesigned based on CUDA to achieve computational speedup. Differential evolution (DE), as a very promising evolutionary algorithm, is highly suitable for parallelization owing to its data parallel algorithmic structure. However, most existing CUDA based DE implementations suffer from excessive low-throughput memory access and less efficient device utilization. This work presents an improved CUDA-based DE to optimize memory and device utilization: several logically-related kernels are combined into one composite kernel to reduce global memory access; kernel execution configuration parameters are automatically determined to maximize device occupancy; streams are employed to enable concurrent kernel execution to maximize device utilization. Experimental results on several numerical problems demonstrate superior computational time efficiency of the proposed method over two recent CUDA-based DE and the sequential DE across varying problem dimensions and algorithmic population sizes.

This work was nominated for the best paper award (finalist) in the Digital Entertainment Technologies and Arts / Parallel Evolutionary Systems session of the Genetic and Evolutionary Computation Conference 2012 (GECCO12) conference [33].

6.4.3. Augmented cumulative distribution networks for multivariate extreme value modelling

Participants: Stéphane Girard, Gildas Mazo, Florence Forbes.

Max-stable distribution functions are theoretically grounded models for modelling multivariate extreme values. However they suffer from some striking limitations when applied to real data analysis due to the intractability of the likelihood when the number of variables becomes high. Cumulative Distribution Networks (CDN's) have been introduced recently in the machine learning community and allow the construction of max-stable distribution functions for which the density can be computed. Unfortunately, we show in this work that the dependence structure expected in the data may not be accurately reflected by max-stable CDN's. To face this limitation, we therefore propose to augment max-stable CDN's with the more standard Gumbel max-stable distribution function in order to enrich the dependence structure [32].

6.4.4. Modelling extremal events

Participants: Stéphane Girard, Jonathan El-Methni, El-Hadji Deme.

Joint work with: Guillou, A. and Gardes, L. (Univ. Strasbourg).

We introduced a new model of tail distributions depending on two parameters $\tau \in [0,1]$ and $\theta > 0$. This model includes very different distribution tail behaviors from Fréchet and Gumbel maximum domains of attraction. In the particular cases of Pareto type tails ($\tau = 1$) or Weibull tails ($\tau = 0$), our estimators coincide with classical ones proposed in the literature, thus permitting us to retrieve their asymptotic normality in an unified way. The first year of the PhD work of Jonathan El-methni has been dedicated to the definition of an estimator of the parameter τ. This permits the construction of new estimators of extreme quantiles. The results are published in [17]. Our future work will consist in proposing a test procedure in order to discriminate between Pareto and Weibull tails.
We are also working on the estimation of the second order parameter ρ (see paragraph 3.3.1). We proposed a new family of estimators encompassing the existing ones (see for instance [64], [63]). This work is in collaboration with El-Hadji Deme, a PhD student from the Université de Saint-Louis (Sénégal). El-Hadji Deme obtained a one-year mobility grant to work within the Mistis team on extreme-value statistics. The results are submitted for publication [49]. We also proposed reduced-bias estimators of the Proportional Hazard Premium for heavy-tailed distributions. The results are submitted for publication [50].

6.4.5. Conditional extremal events

Participants: Stéphane Girard, Gildas Mazo, Jonathan El-methni.

Joint work with: L. Gardes, Amblard, C. (TimB in TIMC laboratory, Univ. Grenoble I) and Daouia, A. (Univ. Toulouse I and Univ. Catholique de Louvain)

The goal of the PhD thesis of Alexandre Lekina was to contribute to the development of theoretical and algorithmic models to tackle conditional extreme value analysis, *ie* the situation where some covariate information X is recorded simultaneously with a quantity of interest Y. In such a case, the tail heaviness of Y depends on X, and thus the tail index as well as the extreme quantiles are also functions of the covariate. We combine nonparametric smoothing techniques [59] with extreme-value methods in order to obtain efficient estimators of the conditional tail index and conditional extreme quantiles. When the covariate is functional and random (random design) and the tail of the distribution is heavy, we focus on kernel methods [18]. We extension to all kind of tails in investigated in [15].

Conditional extremes are studied in climatology where one is interested in how climate change over years might affect extreme temperatures or rainfalls. In this case, the covariate is univariate (time). Bivariate examples include the study of extreme rainfalls as a function of the geographical location. The application part of the study is joint work with the LTHE (Laboratoire d'étude des Transferts en Hydrologie et Environnement) located in Grenoble.

More future work will include the study of multivariate and spatial extreme values. With this aim, a research on some particular copulas [1] has been initiated with Cécile Amblard, since they are the key tool for building multivariate distributions [69]. The PhD theses of Jonathan El-methni and Gildas Mazo should address this issue too.

6.4.6. Level sets estimation

Participant: Stéphane Girard.

Joint work with: Guillou, A. and Gardes, L. (Univ. Strasbourg), Stupfler, G. (Univ. Strasbourg) and Daouia, A. (Univ. Toulouse I and Univ. Catholique de Louvain).

The boundary bounding the set of points is viewed as the larger level set of the points distribution. This is then an extreme quantile curve estimation problem. We proposed estimators based on projection as well as on kernel regression methods applied on the extreme values set, for particular set of points [10].
In collaboration with A. Daouia, we investigate the application of such methods in econometrics [42], [48]: A new characterization of partial boundaries of a free disposal multivariate support is introduced by making use of large quantiles of a simple transformation of the underlying multivariate distribution. Pointwise empirical and smoothed estimators of the full and partial support curves are built as extreme sample and smoothed quantiles. The extreme-value theory holds then automatically for the empirical frontiers and we show that some fundamental properties of extreme order statistics carry over to Nadaraya's estimates of upper quantile-based frontiers.
In the PhD thesis of Gilles Stupfler (co-directed by Armelle Guillou and Stéphane Girard), new estimators of the boundary are introduced. The regression is performed on the whole set of points, the selection of the "highest" points being automatically performed by the introduction of high order moments [19], [20], [21].

6.4.7. Quantifying uncertainties on extreme rainfall estimations

Participant: Stéphane Girard.

Joint work with: Carreau, J. (Hydrosciences Montpellier), Gardes, L. (univ. Strasbourg) and Molinié, G. from Laboratoire d'Etude des Transferts en Hydrologie et Environnement (LTHE), France.

Extreme rainfalls are generally associated with two different precipitation regimes. Extreme cumulated rainfall over 24 hours results from stratiform clouds on which the relief forcing is of primary importance. Extreme rainfall rates are defined as rainfall rates with low probability of occurrence, typically with higher mean return-levels than the maximum observed level. For example Figure 2 presents the return levels for the Cévennes-Vivarais region that can be obtained. It is then of primary importance to study the sensitivity of the extreme rainfall estimation to the estimation method considered.

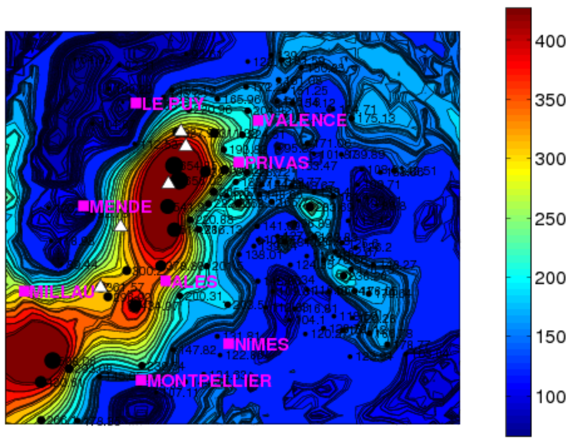


Figure 2. Map of the mean return-levels (in mm) for a period of 10 years.

The obtained results are published in [12].

6.4.8. Retrieval of Mars surface physical properties from OMEGA hyperspectral images.

Participant: Stéphane Girard.

Joint work with: Douté, S. from Laboratoire de Planétologie de Grenoble, France and Saracco, J (University Bordeaux).

Visible and near infrared imaging spectroscopy is one of the key techniques to detect, to map and to characterize mineral and volatile (eg. water-ice) species existing at the surface of planets. Indeed the chemical composition, granularity, texture, physical state, etc. of the materials determine the existence and morphology of the absorption bands. The resulting spectra contain therefore very useful information. Current imaging spectrometers provide data organized as three dimensional hyperspectral images: two spatial dimensions and one spectral dimension. Our goal is to estimate the functional relationship F between some observed spectra and some physical parameters. To this end, a database of synthetic spectra is generated by a physical radiative transfer model and used to estimate F. The high dimension of spectra is reduced by Gaussian regularized sliced inverse regression (GRSIR) to overcome the curse of dimensionality and consequently the sensitivity of the inversion to noise (ill-conditioned problems) [47]. We have also defined an adaptive version of the method which is able to deal with block-wise evolving data streams [46].

6.4.9. Statistical modelling development for low power processor.

Participant: Stéphane Girard.

Joint work with: A. Lombardot and S. Joshi (ST Crolles).

With scaling down technologies to the nanometer regime, the static power dissipation in semiconductor devices is becoming more and more important. Techniques to accurately estimate System On Chip static power dissipation are becoming essential. Traditionally, designers use a standard corner based approach to optimize and check their devices. However, this approach can drastically underestimate or over-estimate process variations impact and leads to important errors.

The need for an effective modeling of process variation for static power analysis has led to the introduction of Statistical static power analysis. Some publication state that it is possible to save up to 50% static power using statistical approach. However, most of the statistical approaches are based on Monte Carlo analysis, and such methods are not suited to large devices. It is thus necessary to develop solutions for large devices integrated in an industrial design flow. Our objective to model the total consumption of the circuit from the probability distribution of consumption of each individual gate. Our preliminary results are published in [23].

MODAL Project-Team

6. New Results

6.1. Model for conditionally correlated categorical data

Participants: Christophe Biernacki, Matthieu Marbac-Lourdelle, Vincent Vandewalle.

An extension of the latent class model is proposed for clustering categorical data by relaxing the classical class conditional independence assumption of variables. In this model, variables are grouped into inter-independent and intra-dependent blocks in order to consider the main intra-class correlations. The dependence between variables grouped into the same block is taken into account by mixing two extreme distributions, which are respectively the independence and the maximum dependence ones. In the conditionally correlated data case, this approach is expected to reduce biases involved by the latent class model and to produce a meaningful model with few additional parameters. The parameters estimation by maximum likelihood is performed by an EM algorithm while a MCMC algorithm avoiding combinatorial problems involved by the block structure search is used for model selection. Applications on sociological and biological data sets bring out the proposed model interest. These results strengthen the idea that the proposed model is meaningful and that biases induced by the conditional independence assumption of the latent class model are reduced. This model was used in September for software components data set of Philippe Merle (ADAM Team Inria Lille).

A conference paper [26] and a poster workshop [35] have been presented. A preprint has been also written [45]. Furthermore, an R package is currently under development.

6.2. Model-based clustering for multivariate partial ranking data

Participants: Christophe Biernacki, Julien Jacques.

The first model-based clustering algorithm dedicated to multivariate partial ranking data has been developed in [43]. This is an extension of the (ISR) model for ranking data published in [4]. The proposed algorithm has allowed to exhibit regional alliances between European countries in the Eurovision contest, which are often suspected but never proved.

6.3. A new probability distribution for ordinal data

Participants: Christophe Biernacki, Julien Jacques.

In [21], a probability distribution for univariate ordinal data is proposed from a stochastic dichotomic search algorithm in a sorting table. Interest of this approach is to give a specific model for ordinal data, without any reference to numerical or nominal data, as it is often the case. The resulting distribution is governed by a position and a dispersion parameter, and is easily estimated by using an EM algorithm.

6.4. Clustering and variable selection in regression

Participants: Christophe Biernacki, Julien Jacques, Loic Yengo.

The works presented in [28] address the issue of simultaneous linear regression and clustering of predictors. A new framework is proposed that both sidesteps optimization challenges and improves prediction performance. In that framework, regression coefficients are assumed to be drawn from a gaussian mixture distribution. Prediction is thus performed using the conditional distribution of the regression coefficients given the data, while clusters are easily derived from posterior distribution in groups given the data.

6.5. Mixture of Gaussians with Missing Data

Participants: Christophe Biernacki, Vincent Vandewalle.

The generative models allow to handle with missing data. This can be easily performed by using the EM algorithm, which has a closed form M-step in the Gaussian setting. This can for instance be useful for distance estimation with missing data. It has been proposed in [18] to improve the distance estimation by fitting a mixture of Gaussian distribution instead of a considering only one Gaussian component. An extension of the previous work including the high setting has been submitted in Neurocomputing journal. This is a joined work with Emil Eirola and Amaury Lendrasse .

A parallel work is in progress on the mixture degeneracy when considering mixture of Gaussians with missing data. It have been experimentally noticed that the degeneracy in this case is particularly slow. This behaviour is different from the usual setting of degeneracy with mixture of Gaussians which is usually rather fast. We are working on the theoretical characterization of this behaviour around a degenerated solution.

6.6. Transfer learning in model-based clustering

Participant: Christophe Biernacki.

In many situations one needs to cluster several datasets, possibly arising from different populations, instead of a single one, into partitions with identical meaning and described by similar features. Such situations involve commonly two kinds of standard clustering processes. The samples are clustered traditionally either as if all units arose from the same distribution, or on the contrary as if the samples came from distinct and unrelated populations. But a third situation should be considered: As the datasets share statistical units of same nature and as they are described by features of same meaning, there may exist some link between the samples. We propose a linear stochastic link between the samples, what can be justified from some simple but realistic assumptions, both in the Gaussian and in the t mixture model-based clustering context [37]. This is a joint work with Alexandre Lourme.

A book chapter about transfer learning (including clustering, classification and regression) has been also published [37]. It is a joint work with Farid Beninel, Charles Bouveyron, Julien Jacques and Alexandre Lourme.

6.7. Gaussian Models Scale Invariant and Stable by Projection

Participant: Christophe Biernacki.

Gaussian mixture model-based clustering is now a standard tool to determine an hypothetical underlying structure into continuous data. However many usual parsimonious models, despite their appealing geometrical interpretation, suffer from major drawbacks as scale dependence or unsustainability of the constraints by projection. In this work we present a new family of parsimonious Gaussian models based on a variance-correlation decomposition of the covariance matrices. These new models are stable by projection into the canonical planes and, so, faithfully representable in low dimension. They are also stable by modification of the measurement units of the data and such a modification does not change the model selection based on likelihood criteria. We highlight all these stability properties by a specific geometrical representation of each model. A detailed GEM algorithm is also provided for every model inference. Then, on biological and geological data, we compare our stable models to standard geometrical ones.

This work is was presented as a poster to workshop [31] and is also a preprint [41] currently in revision in an international journal. This is a joint work with Alexandre Lourme.

6.8. Decorrelating variables in high dimension for linear regression

Participants: Christophe Biernacki, Clément Thery.

Databases from the steel industry are often large (very long process with many parameters) and have strong correlations between variables. Some variables may be written directly in terms of other via physical models or related by definition. Moreover the process, which is specific to the type of finished product, conditions most of the process parameters and therefore induces strong correlations between variables. The main idea is to consider some form of sub-regressions models, some variables defining others. We can then remove temporarily some of the variables to overcome ill-conditioned matrices inherent in linear regression and then reinject the deleted information, based on the struc- ture that links the variables. The final model therefore takes into account all the variables but without suffering from the consequences of correlations between variables or high dimension. This research is placed in a steel industry context (Arcelor-Mittal Dunkerque).

The work has been presented to a conference [27] and as a poster to a workshop [36]. It is a joint work with Gaétan Loridant from Arcelor-Mittal.

6.9. Model-based clustering for multivariate functional data

Participants: Julien Jacques, Cristian Preda.

We developed in [19] an extension of the model-based clustering algorithm for univariate functional data proposed in [20], [23], [11] to the case of multivariate functional data. For this, multivariate functional principal components analysis is defined and a parametric mixture model is proposed and estimated by an EM-like algorithm. Results on simulated and real datasets have shown the efficiency of the proposed method.

6.10. A method to combine combinatorial optimization and statistics to mine high-throughput genotyping data

Participants: Julie Hamon, Julien Jacques, Clarisse Dhaenens.

In the context of genomic analysis (collaboration with Genes Diffusion), dealing with high-throughput genotyping data, the objective of our study is to select a subset of SNPs (single nucleotide polymorphisms) explaining a trait of interest. We propose in [33] and [32] a method combining combinatorial optimization and statistics to extract a subset of interesting SNPs. The combinatorial part aims at exploring in a efficient way the large search space induced by the large number of possible subsets and statistics are used to evaluate the selection. We propose a first method based on an ILS (iterated local search) and using a regression. Three criteria used to evaluate the quality of the regression are compared. One of them (the k-fold validation) shows better performance. We also compare this approach to classical statistical approaches on simulated datasets. Results are promising as the proposed approach outperforms most of these statistical approaches.

6.11. Wavelet based clustering using mixed effects functional models

Participant: Guillemette Marot.

Curve clustering in the presence of inter-individual variability has longly been studied, especially using splines to account for functional random effects. However splines are not appropriate when dealing with high-dimensional data and can not be used to model irregular curves such as peak-like data. We propose a wavelet based clustering procedure ([6]) and apply it to high dimensional data. We suggest a dimension reduction step based on wavelet thresholding adapted to multiple curves and using an appropriate structure for the random effect variance, we ensure that both fixed and random effects lie in the same functional space even when dealing with irregular functions that belong to Besov spaces. In the wavelet domain, our model resumes to a linear mixed-effects model that can be used for a model-based clustering algorithm and for which we develop an EM- algorithm for maximum likelihood estimation. An R package curvclust implementing this procedure has been posted this year to the CRAN, the official website of the R software.

6.12. Comparison of normalisation procedures in RNA-sequencing before differential analysis

Participant: Guillemette Marot.

The continuing technical improvements and decreasing cost of next-generation sequencing technologies have made RNA sequencing (RNA-seq) a popular choice for gene expression studies. Several methods for the normalization of RNAseq data (removal of errors due to the small number of samples, corrections for sequence composition) have been proposed in recent years. With the Statomique Consortium, we have compared seven normalisation methods, discarded two out of them (although still widely used). We give practical recommendations on the appropriate normalization method to be used and its impact on the differential analysis of RNA-seq data in the paper ([14]).

6.13. Change point detection algorithm

Participant: Alain Célisse.

We develop a new change-point detection algorithm where focus is given to detect changes in the whole distribution of data. This challenging problem is addressed by use of kernels which enable us to deal with non-vectorial data of aby type (graphs, DNA sequences, etc). A preprint has been submitted ([46]).

6.14. Cross validation algorithms

Participant: Alain Célisse.

The performance of Cross-validation (CV) algorithms are assessed for estimating the risk as well as for model selection. Whereas optimality of leave-one-out (LOO) cross-validation is proved for risk estimation, it is no longer the case for model selection. In the latter setup, conditions are derived that lead to optimality for leave-p-cross-validation (LPO) when p is larger than 1. See for details [47].

6.15. Stochastic Block Model

Participant: Alain Célisse.

The convergence of maximum likelihood and variational estimators in a random graph model called Stochastic Block model is addressed. To the best of our knowledge, these are the first results providing consistency for maximum likelihood and variational estimators in that model. See [5].

6.16. Approximations for scan statistics.

Participants: Alexandru Amarioarei, Cristian Preda.

Accurate approximations for the distribution of extremes of 1-dependent stationary sequences are developed (see [38]). Viewed as maximum of some particular sequence of 1-dependent random variables, we provide sharp error bounds and approximations for the distribution of the three-dimensional scan statistics (see [39]). The Binomial and Poisson models are considered.

NACHOS Project-Team

6. New Results

6.1. Discontinuous Galerkin methods for Maxwell's equations

6.1.1. DGTD-$\mathbb{P}_p$ method based on hierarchical polynomial interpolation

Participants: Loula Fezoui, Stéphane Lanteri.

The DGTD (Discontinuous Galerkin Time Domain) method originally proposed by the team for the solution of the time domain Maxwell's equations [14] relies on an arbitrary high order polynomial interpolation of the component of the electromagnetic field, and its computer implementation makes use of nodal (Lagrange) basis expansions on simplicial elements. The resulting method is often denoted by DGTD-$\mathbb{P}_p$ where p refers to the interpolation degree that can be defined locally i.e. at the element level. In view of the design of a hp-adaptive DGTD method, i.e. a solution strategy allowing an automatic adaptation of the interpolation degree p and the discretization step h, we now investigate alternative polynomial interpolation and in particular those which lead to hierarchical or/and orthogonal basis expansions. Such basis expansions on simplicial elements have been extensively studied in the context of continuous finite element formulations (e.g. [52]) and have thus been designed with global conformity requirements (i.e. H_1, $H(\text{rot})$ or (div)) whose role in the context of a discontinuous Galerkin formulation has to be clarified. This represents one of the objectives of this study.

6.1.2. DGTD-$\mathbb{P}_p\mathbb{Q}_k$ method on multi-element meshes

Participants: Clément Durochat, Stéphane Lanteri, Raphael Léger, Claire Scheid, Mark Loriot [Distene, Pôle Teratec, Bruyères-le-Chatel].

In this work, we study a multi-element DGTD method formulated on a hybrid mesh which combines a structured (orthogonal) discretization of the regular zones of the computational domain with an unstructured discretization of the irregularly shaped objects. The general objective is to enhance the flexibility and the efficiency of DGTD methods for large-scale time domain electromagnetic wave propagation problems with regards to the discretization process of complex propagation scenes. With this objective in mind, we have designed and analyzed a DGTD-$\mathbb{P}_p\mathbb{Q}_k$ method formulated on non-conforming hybrid quadrangular/triangular meshes (2D case) or non-conforming hexahedral/tetrahedral meshes (3D case) for the solution of the time domain Maxwell's equations.

6.1.3. DGTD-$\mathbb{P}_p$ method for Debye media and applications to biolectromagnetics

Participants: Claire Scheid, Maciej Klemm [Communication Systems & Networks Laboratory, Centre for Communications Research, University of Bristol, UK], Stéphane Lanteri.

This work is undertaken in the context of a collaboration with the Communication Systems & Networks Laboratory, Centre for Communications Research, University of Bristol (UK). This laboratory is studying imaging modalities based on microwaves with applications to dynamic imaging of the brain activity (Dynamic Microwave Imaging) on one hand, and to cancerology (imaging of breast tumors) on the other hand. The design of imaging systems for these applications is extensively based on computer simulation, in particular to assess the performances of the antenna arrays which are at the heart of these systems. In practice, one has to model the propagation of electromagnetic waves emitted from complex sources and which propagate and interact with biological tissues. In relation with these issues, we study the extension of the DGTD-$\mathbb{P}_p$ method originally proposed in [14] to the numerical treatment of electromagnetic wave propagation in dispersive media. We consider an approach based on an auxiliary differential equation modeling the time evolution of the electric polarization for a dispersive medium of Debye type (other dispersive media will be considered subsequently). The stability and a priori convergence analysis of the resulting DGTD-$\mathbb{P}_p$ method has been recently studied [25], and its application to the simulation of the propagation in realistic geometrical models of head tissues is underway.

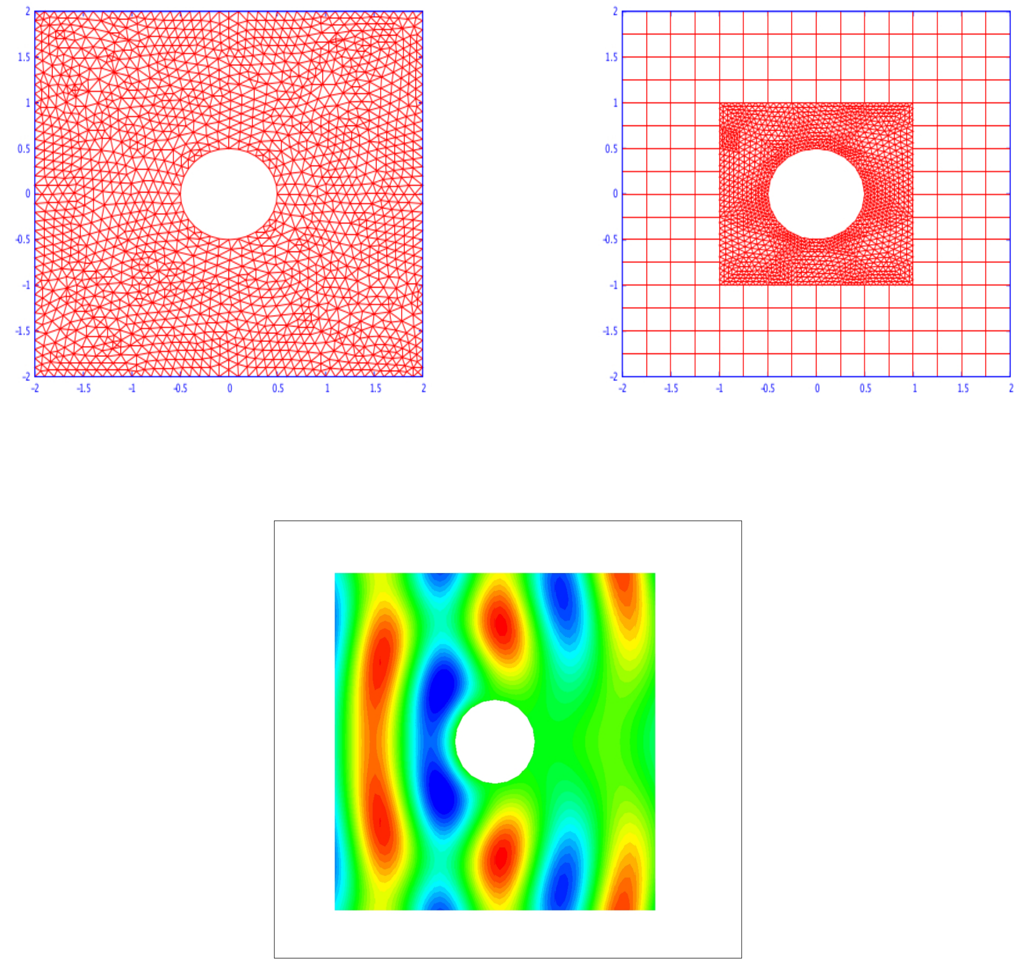


Figure 2. Scattering of a plane wave by an disk. Conforming triangular mesh (top left) and non-conforming quadrangular/triangular mesh (top right). Contour lines of electrical field component E_z from a simulation with a DGTD-$\mathbb{P}_2\mathbb{Q}_4$ method (bottom).

6.1.4. DGTD-$\mathbb{P}_p$ method for Drude media and applications to nanophotonics

Participants: Claire Scheid, Maciej Klemm [Communication Systems & Networks Laboratory, Centre for Communications Research, University of Bristol, UK], Stéphane Lanteri, Jonathan Viquerat.

Nanostructuring of materials has opened up a number of new possibilities for manipulating and enhancing light-matter interactions, thereby improving fundamental device properties. Low-dimensional semiconductors, like quantum dots, enable one to catch the electrons and control the electronic properties of a material, while photonic crystal structures allow to synthesize the electromagnetic properties. These technologies may, e.g., be employed to make smaller and better lasers, sources that generate only one photon at a time, for applications in quantum information technology, or miniature sensors with high sensitivity. The incorporation of metallic structures into the medium allows one to exploit plasmonic effects and adds further possibilities for manipulating the propagation of electromagnetic waves. In particular, this allows subwavelength localisation of the electromagnetic field and, by subwavelength structuring of the material, novel effects like negative refraction, e.g. enabling super lenses, may be realized. Nanophotonics is the recently emerged, but already well defined, field of science and technology aimed at establishing and using the peculiar properties of light and light-matter interaction in various nanostructures. Because of its numerous scientific and technological applications (e.g. in relation to telecommunication, energy production and biomedicine), nanophotonics represents an active field of research increasingly relying on numerical modeling beside experimental studies. We have started this year a new research direction aiming at the numerical modeling of electromagnetic wave interaction with nanoscale metallic structures. In this context, one has to take into account the dispersive characteristics of cartain metals in the frequency range of interest to nanophotonics. As a first step in this direction, we have considered an auxiliary differential equation approach for the numerical treatment of a Drude dispersion model in the framework of a DGFD-$\mathbb{P}_p$ method.

6.1.5. Frequency domain hybridized DGFD-$\mathbb{P}_p$ methods

Participants: Stéphane Lanteri, Liang Li [Faculty Member, School of Mathematical Sciences, Institute of Computational Science, University of Electronic Science and Technology of China Chengdu, China], Ronan Perrussel [Laplace Laboratory, INP/ENSEEIHT/UPS, Toulouse].

For certain types of problems, a time harmonic evolution can be assumed leading to the formulation of the frequency domain Maxwell equations, and solving these equations may be more efficient than considering the time domain variant. We are studying a high order Discontinuous Galerkin Frequency Domain (DGFD-$\mathbb{P}_p$) method formulated on unstructured meshes for solving the 2D and 3D time harmonic Maxwell equations. However, one major drawback of DG methods is their intrinsic cost due to the very large number of globally coupled degrees of freedom as compared to classical high order conforming finite element methods. Different attempts have been made in the recent past to improve this situation and one promising strategy has been recently proposed by Cockburn *et al.* [47] in the form of so-called hybridizable DG formulations. The distinctive feature of these methods is that the only globally coupled degrees of freedom are those of an approximation of the solution defined only on the boundaries of the elements. This work is concerned with the study of such Hybridizable Discontinuous Galerkin (HDG) methods for the solution of the system of Maxwell equations in the time domain when the time integration relies on an implicit scheme, or in the frequency domain. In particular, we have recently designed a HDGFD-$\mathbb{P}_p$ method for the solution of the 2D frequency domain Maxwell equations [22] and, based on the very promising results obtained in this study, the extension to the more challenging 3D case has been initiated.

6.1.6. Exact transparent condition in a DGFD-$\mathbb{P}_p$ method

Participants: Mohamed El Bouajaji, Nabil Gmati [ENIT-LAMSIN, Tunisia], Stéphane Lanteri, Jamil Salhi [ENIT-LAMSIN, Tunisia].

In the numerical treatment of propagation problems theoretically posed in unbounded domains, an artificial boundary is introduced on which an absorbing condition is imposed. For the frequency domain Maxwell equations, one generally use the Silver-Müller condition which is a first order approximation of the exact radiation condition. Then, the accuracy of the numerical treatment greatly depends on the position of the artificial boundary with regards to the scattering object. In this work, we have conducted a preliminary study aiming at improving this situation by using an exact transparent condition in place of the Silver-Müller condition. Promising results have been obtained in the 2D case [26].

6.2. Discontinuous Galerkin methods for the elastodynamic equations

6.2.1. DGTD-$\mathbb{P}_p$ method for viscoelastic media

Participants: Nathalie Glinsky, Stéphane Lanteri, Fabien Peyrusse.

We continue developing high order non-dissipative discontinuous Galerkin methods on simplicial meshes for the numerical solution of the first order hyperbolic linear system of elastodynamic equations. These methods share some ingredients of the DGTD-$\mathbb{P}_p$ methods developed by the team for the time domain Maxwell equations among which, the use of nodal polynomial (Lagrange type) basis functions, a second order leap-frog time integration scheme and a centered scheme for the evaluation of the numerical flux at the interface between neighboring elements. The resulting DGTD-$\mathbb{P}_p$ methods have been validated and evaluated in detail in the context of propagation problems in both homogeneous and heterogeneous media including problems for which analytical solutions can be computed. Particular attention was given to the study of the mathematical properties of these schemes such as stability, convergence and numerical dispersion.

A recent novel contribution is the extension of the DGTD method to include viscoelastic attenuation. For this, the velocity-stress first-order hyperbolic system is completed by additional equations for the anelastic functions including the strain history of the material. These additional equations result from the rheological model of the generalized Maxwell body and permit the incorporation of realistic attenuation properties of viscoelastic material accounting for the behaviour of elastic solids and viscous fluids. In practice, we need solving 3L additional equations in 2D (and 6L in 3D), where L is the number of relaxation mechanisms of the generalized Maxwell body. This method has been implemented in 2D and validated by comparison to results obtained by a finite-difference method, in particular for wave propagation in a realistic basin of the area of Nice (south of France) [40]-[35].

6.2.2. DGTD-$\mathbb{P}_p$ method for the assessment of topographic effects

Participants: Etienne Bertrand [CETE Méditerranée], Nathalie Glinsky.

This study addresses the numerical assessment of site effects especially topographic effects. The study of measurements and experimental records proved that seismic waves can be amplified at some particular locations of a topography. Numerical simulations are exploited here to understand further and explain this phenomenon. The DGTD-$\mathbb{P}_p$ method has been applied to a realistic topography of Rognes area (where the Provence earthquake occured in 1909) to model the observed amplification and the associated frequency. Moreover, the results obtained on several homogeneous and heterogeneous configurations prove the influence of the medium in-depth geometry on the amplifications measures at the surface [38].

6.2.3. DGTD-$\mathbb{P}_p$ method for arbitrary heterogeneous media

Participants: Nathalie Glinsky, Diego Mercerat [CETE Méditerranée].

We have recently devised an extension of the DGTD method for elastic wave propagation in arbitrary heterogeneous media. In realistic geological media (sedimentary basins for example), one has to include strong variations in the material properties. Then, the classical hypothesis that these properties are constant within each element of the mesh can be a severe limitation of the method, since we need to discretize the medium with very fine meshes resulting in very small time steps. For these reasons, we propose an improvement of the DGTD method allowing non-constant material properties within the mesh elements. A change of variables on the stress components allows writing the elastodynamic system in a pseudo-conservative form. Then, the introduction of non-constant material properties inside an element is simply treated by the calculation, via convenient quadrature formulae, of a modified local mass matrix depending on these properties. This new extension has been validated for a smoothly varying medium or a strong jump between two media, which can be accurately approximated by the method, independently of the mesh [39].

6.2.4. DGFD-$\mathbb{P}_p$ method for frequency domain elastodynamics

Participants: Hélène Barucq [MAGIQUE3D project-team, Inria Bordeaux - Sud-Ouest], Marie Bonnasse, Julien Diaz [MAGIQUE3D project-team, Inria Bordeaux - Sud-Ouest], Stéphane Lanteri.

We have started this year a research direction aiming at the development of high order discontinuous Galerkin methods on unstructured meshes for the simulation of frequency domain elastodynamic and viscelastic wave propagation. This study is part of the Depth Imaging Partnership (DIP) between Inria and TOTAL. The PhD thesis of Marie Bonnasse is at the heart of this study which is funded by TOTAL.

6.3. Time integration strategies and resolution algorithms

6.3.1. Hybrid explicit-implicit DGTD-$\mathbb{P}_p$ method

Participants: Stéphane Descombes, Stéphane Lanteri, Ludovic Moya.

Existing numerical methods for the solution of the time domain Maxwell equations often rely on explicit time integration schemes and are therefore constrained by a stability condition that can be very restrictive on highly refined meshes. An implicit time integration scheme is a natural way to obtain a time domain method which is unconditionally stable. Starting from the explicit, non-dissipative, DGTD-$\mathbb{P}_p$ method introduced in [14], we have proposed the use of Crank-Nicolson scheme in place of the explicit leap-frog scheme adopted in this method [4]. As a result, we obtain an unconditionally stable, non-dissipative, implicit DGTD-$\mathbb{P}_p$ method, but at the expense of the inversion of a global linear system at each time step, thus obliterating one of the attractive features of discontinuous Galerkin formulations. A more viable approach for 3D simulations consists in applying an implicit time integration scheme locally i.e in the refined regions of the mesh, while preserving an explicit time scheme in the complementary part, resulting in an hybrid explicit-implicit (or locally implicit) time integration strategy. In [6], we conducted a preliminary numerical study of a hyrbid explicit-implicit DGTD-$\mathbb{P}_p$ method, combining a leap-frog scheme and a Crank-Nicolson scheme, and obtained promising results. More recently, we further investigated two such strategies, both theoretically (especially, convergence in the ODE and PDE senses) [24] and numerically in the 2D case [23]. A last topic is to propose higher order time integration techniques based on the second-order locally implicit method to fully exploit the attractive features of this approach combined with a DG discretisation which allows to easily increase the spatial convergence order. Promising results in 2D reaching high order in time, between $3, 5$ and 4, have been obtained in [33] by applying Richardson extrapolation and composition methods.

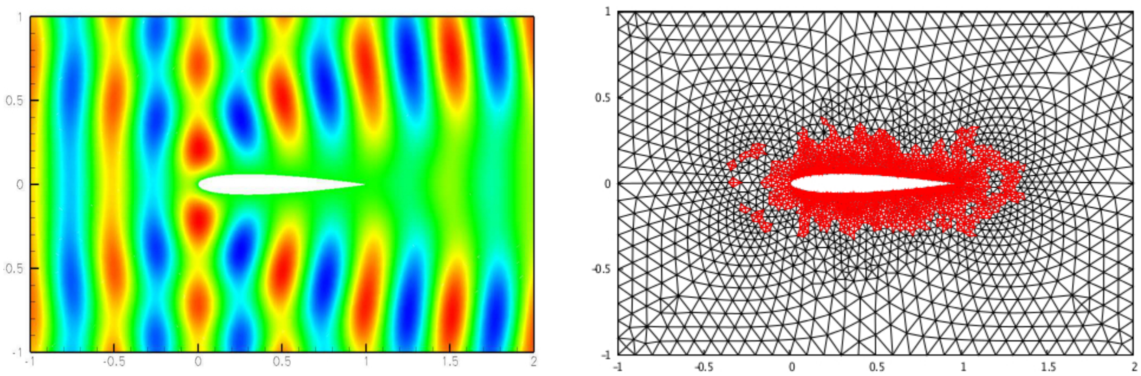


Figure 3. Scattering of a plane wave by an airfoil profile. Contour lines of electrical field component E_z (left) and locally refined triangular mesh with partitioning in explicit/implicit zones (right).

6.3.2. Optimized Schwarz algorithms for the frequency domain Maxwell equations

Participants: Victorita Dolean, Mohamed El Bouajaji, Martin Gander [Mathematics Section, University of Geneva], Stéphane Lanteri, Ronan Perrussel [Laplace Laboratory, INP/ENSEEIHT/UPS, Toulouse].

Even if they have been introduced for the first time two centuries ago, over the last two decades, classical Schwarz methods have regained a lot of popularity with the developement of parallel computers. First developed for the elliptic problems, they have been recently extended to systems of hyperbolic partial differential equations, and it was observed that the classical Schwartz method can be convergent even without overlap in certain cases. This is in strong contrast to the behavior of classical Schwarz methods applied to elliptic problems, for which overlap is essential for convergence. Over the last decade, optimized versions of Schwarz methods have been developed for elliptic partial differential equations. These methods use more effective transmission conditions between subdomains, and are also convergent without overlap for elliptic problems. The extension of such methods to systems of equations and more precisely to Maxwell's system (time harmonic and time discretized equations) has been studied in [8]. The optimized interface conditions proposed in [8] were devised for the case of non-conducting propagation media. We have recently studied the formulation of such conditions for conducting media [17]. Besides, we have also proposed an appropriate discretization strategy of these optimized Schwarz algorithms in the context of a high order DGFD-$\mathbb{P}_p$ method formulated on unstructured triangular meshes for the solution of the 2D frequency domain Maxwell equations [28].

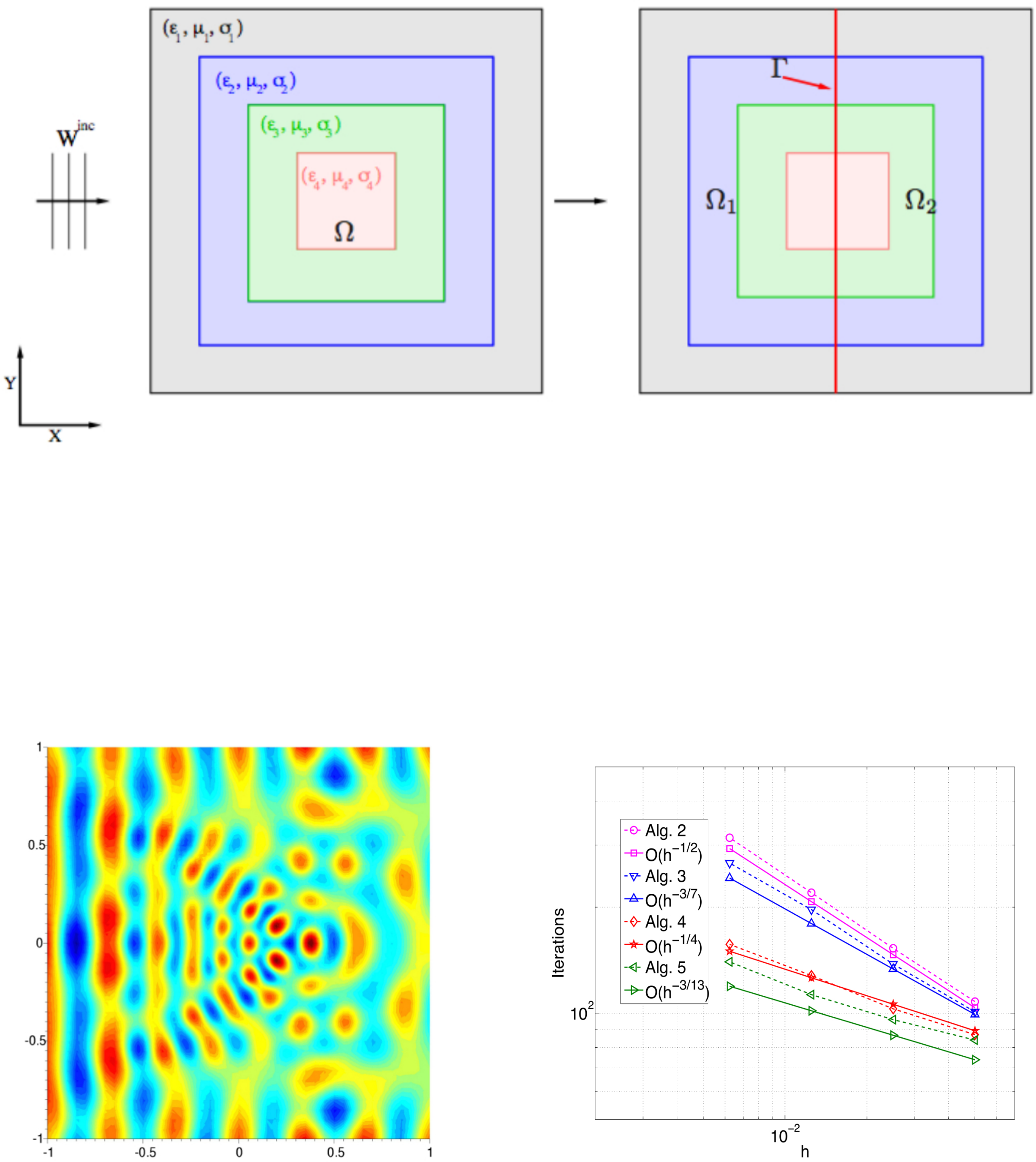


Figure 4. Propagation of a plane wave in a multilayered heterogeneous medium. Problem setting and two-subdomain decompositin (top). Contour lines of the real part of the E_z component of the electrical field (bottom left) and asymptotic convergence of the optimized Schwarz algorithms (bottom right).

NANO-D Team

6. New Results

6.1. Adaptively Restrained Particle Simulations

Participants: Svetlana Artemova, Stephane Redon.

Last year, we have introduced a novel, general approach to speed up particle simulations that we call Adaptively Restrained Particle Simulations (ARPS). This year we continued working on this approach. The obtained results have been published in Physical Review Letters [3], and the patent describing the theoretical basis and the algorithms for the numerical realization of ARPS has been deposited.

Particle simulations are widely used in physics, chemistry, biology [13], [14], and even computer graphics [9], and faster simulations (in particular ARPS) may result in progress on many challenging problems, e.g., protein folding, diffusion across bio-membranes, fracture in metals, ion implantation, etc.

ARPS rely on an adaptively restrained (AR) Hamiltonian used to describe a system of N particles:

$$H_{AR}(\mathbf{q},\mathbf{p}) = \frac{1}{2}\mathbf{p}^T\Phi(\mathbf{q},\mathbf{p})\mathbf{p} + V(\mathbf{q}).$$

This Hamiltonian has an unusual inverse inertia matrix $\Phi(\mathbf{q},\mathbf{p})$, which is made a general function of phase-space coordinates. The precise form of this matrix can be chosen according to the system under study and the problem stated.

We have proposed a particular (diagonal) form of the inverse inertia matrix for the simulations in Cartesian coordinates. In this case, Φ adaptively switches on and off positional degrees of freedom of individual particles while letting particle momenta evolve. The decision whether the particle is restrained or not depends on the particle's momentum, and, precisely, on it's kinetic energy. Two user-defined thresholds regulate the amount of simplification of the particle's motion. When a module of a particle's momentum becomes small enough (without necessarily becoming zero), the particle completelystops moving. Even when a particle is fully restrained, though, its momentum may continue to change, and its kinetic energy might become large enough again for the particle to resume moving. In general, ARPS restrain and release particles repeatedly over time.

This approach has numerous advantages: (a) it is mathematically grounded and is able to produce long, stable simulations; (b) it does not require modifications to the simulated interaction potential, so that any suitable existing force-field can be directly used with ARPS; (c) under frequently-used assumptions on the interaction potential, ARPS make it possible to reduce the number of forces that have to be updated at each time step, which may significantly speed up simulations; (d) when performing constant-energy simulations, ARPS allow users to finely and continuously trade between precision and computational cost, andrapidly obtain approximate trajectories; (e) the trade-off between precision and cost may be chosen for each particle independently, so that users may arbitrarily focus ARPS on specific regions of the simulated system (e.g., a polymer in a solvent); (f) most important, when performing Adaptively Restrained Molecular Dynamics (ARMD) in the canonical (NVT) ensemble, correct static equilibrium properties can be computed.

We have demonstrated the advantages of ARPS on several numerical experiments. For example, a planar collision cascade study in Fig. 7 shows how ARPS make it possible to smoothly trade between precision and speed of the simulation. Reference simulations were derived from the usual Hamiltonian $H(\mathbf{q},\mathbf{p}) = \frac{1}{2}\,\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} + V(\mathbf{q})$.

6.2. Hierarchical Adaptively Restrained Particle Simulations

Participants: Svetlana Artemova, Stephane Redon.

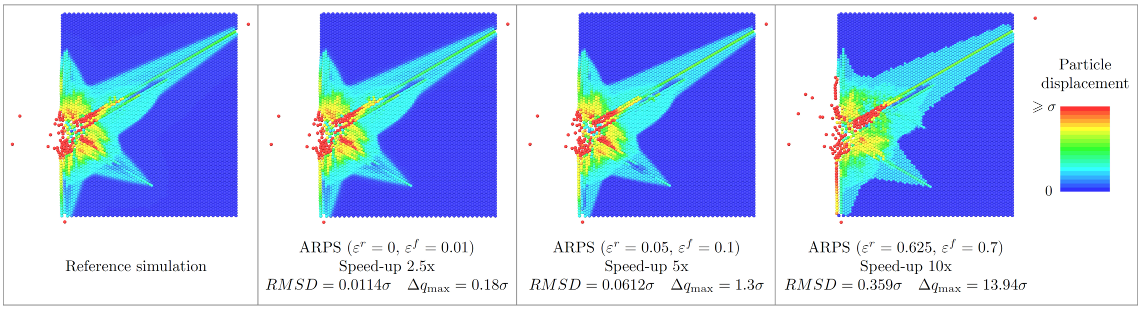


Figure 7. Simulating a collision cascade with controlled precision. Adaptively restrained simulations allow us to smoothly trade between precision and speed. Even for large speed-ups (up to 10x) the features of the shock are extremely well preserved.

It has been shown that algorithms relying on hierarchical representations of molecular systems may accelerate molecular simulations: for example, divide-and-conquer approach for simulations in internal coordinates [10], [11], adaptive algorithms for dynamics of articulated bodies [15], algorithms for neighbor search for system with symmetries [12] or for large rigid molecules [8].

Therefore, we were interested in combining hierarchically-based algorithms with Adaptively Restrained Particle Simulations (ARPS). Precisely, as with classical ARPS, we have considered the adaptively restrained (AR) Hamiltonian:

$$H_{AR}(\mathbf{q}, \mathbf{p}) = \frac{1}{2}\mathbf{p}^T \Phi(\mathbf{q}, \mathbf{p})\mathbf{p} + V(\mathbf{q}),$$

but we have introduced a different form of the inverse inertia matrix $\Phi(\mathbf{q}, \mathbf{p})$. In this case, again, positional degrees of freedom are adaptively switched on and off during the simulation, but, these are *relative* positional degrees of freedom in the system, and not the positional degrees of freedom of individual particles. Precisely, particles are grouped together into rigid bodies according to the tree representation and released repeatedly during the simulation. We call this approach hierarchical Adaptively Restrained Particle Simulations (hierarchical ARPS).

We have performed several numerical experiments to illustrate this new approach. For example, in Fig. 8 we present the planar collision cascade study.

For hierarchical AR simulations, obtained results depend on the tree representation of the system: for the results demonstrated in Fig. 8 the tree was constructed in a top-down manner by recursive dividing of the system in halves and, therefore, the squares of different levels are being activated by the shock.

To clearly demonstrate the effect of the tree, we provide the results for the same four simulations with another tree built in a bottom-up manner by grouping the particles pairwise according to their sequence number (they were enumerated, first, along the y-axis, vertically, and then, along the x-axis, horizontally). These results are shown in Fig. 9 , and are rather different from those in Fig. 8 : vertical lines are being activated when the central part of the plane is reached by the shock.

The patent reporting the principles and the algorithms used to implement hierarchical ARPS has been deposited.

6.3. Interactive quantum chemistry

Participants: Mael Bosson, Caroline Richard, Antoine Plet, Sergei Grudinin, Stephane Redon.

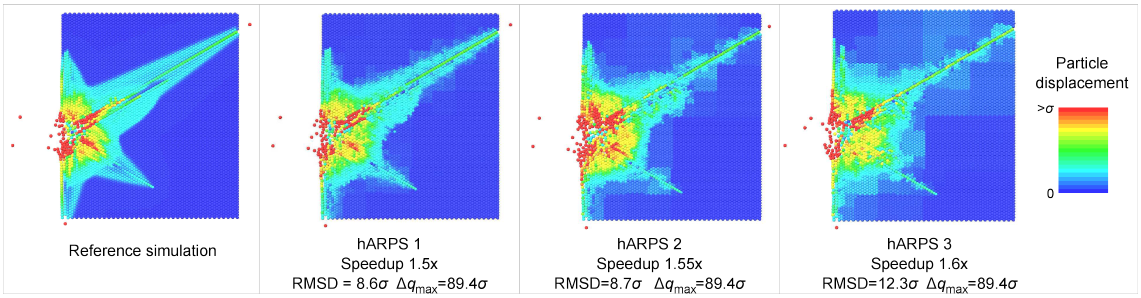


Figure 8. Simulating a collision cascade with controlled precision. Hierarchical adaptively restrained simulations allow us to smoothly trade between precision and speed. The main features of the shock are preserved. The binary tree representation was constructed top-down.

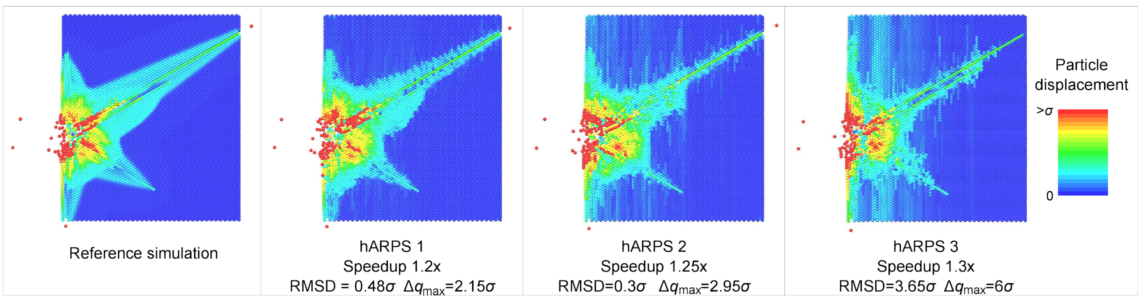


Figure 9. Simulating a collision cascade with controlled precision. Hierarchical adaptively restrained simulations allow us to smoothly trade between precision and speed. The main features of the shock are preserved. The binary tree representation was constructed bottom-up.

Interactive simulation tools allow users to take advantage of their knowledge and intuition to understand physical properties and prototype new devices. To accurately describe bond breaking, bond formation, charge transfer or other electronic phenomena, interactive simulation should ideally be based on quantum mechanics. However, solving quantum chemistry models at interactive rates is a challenging task. Thanks to the algorithms developed in the group, SAMSON is the first software to propose interactive quantum chemistry.

A first contribution allows for interactive quantum chemistry with systems up to a few hundred atoms [6]. The method is based on a divide-and-conquer (D&C) approach. The D&C technique subdivides the system into many subsystems (a–h on the Figure 10). Each of them involves a diagonalisation at each time step. To treat larger systems, we introduce a new algorithm: Block-Adaptive Quantum Mechanics (BAQM) [5] from the combination of two new components.

- **Block-adaptive Cartesian mechanics**

  By freezing atomic positions in some subsystems (d–h on the Figure 10) (with atoms in blue), we may avoid updating some eigenproblems. The Block-adaptive Cartesian mechanics component takes advantage of this to control the simulation cost by adaptively adjusting the number of diagonalisations, based on the forces applied to the atoms. Only the subsystems with the largest applied forces are allowed to have mobile atoms.

- **Adaptive reduced-basis quantum mechanics**

  Solving even just one of the subsystem's eigenproblem may be too costly to achieve interactive rates. The Adaptive reduced-basis quantum mechanics component projects the equation in an adaptive reduced basis composed of low-energy eigenvectors that have been computed at a previous time step, to benefit from temporal coherence between successive eigenproblems (subsystems (b) and (c) with atoms in black and white on the Figure 10). We use a simple distance to decide on the fly when to automatically update the reduced basis during the simulation (subsystem (a) with atoms in red on the Figure 10).

We demonstrated that BAQM may accelerate geometry optimization for several atomic systems. Indeed, each step is solved significantly faster by constraining some nuclei and electrons, and, by focusing computational resources on the most active parts of the system, we obtain a faster potential energy descent. The proposed BAQM approach also allows for interactive rates with many atomic systems.

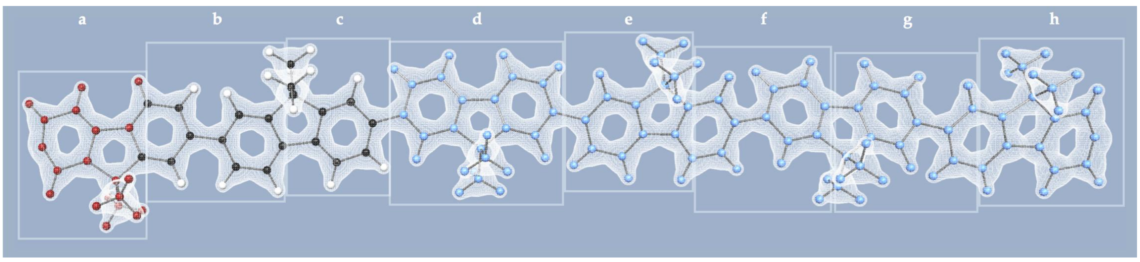


Figure 10. ***Interactive editing of a polyflurorene molecule with the BAQM algorithm***

6.4. Molecular Docking

6.4.1. Development of a new Knowledge-Based Potential for Protein-Ligand Interactions

Participants: Sergei Grudinin, Georgy Cheremovskiy.

Macromolecular complexes formed by proteins with small molecules (ligands) play an important role in many biological processes such as signal transduction, cell regulation, etc. Experimental methods for determining the structures of molecular complexes have a very high cost and still involve many difficulties. Therefore, computational methods, such as molecular docking, are typically used for predicting binding modes and affinities, which are essential to understand molecular interaction mechanisms and design new drugs.

Databases containing three-dimensional protein-ligand structures determined by experimental techniques grow very rapidly. In 2011, the PDB (Protein Data Bank) contained about 70,000 of protein structures, with almost 8,000 structures of protein-ligand complexes having refined binding affinity data. The CSD (Cambridge Structural Database), a database for small molecules, contained about 500,000 entries at the beginning of 2012. Thus, we believe that computational tools based on statistical information extracted from three-dimensional structures of protein-ligand complexes will play an ever more increasing role in the functional study of proteins as well as in structure-based drug design and other fields.

We proposed and validated a new statistical method that predicts binding modes and affinities of protein-ligand complexes. To do so, we have developed a novel machine-learning-based approach. Precisely, we have formulated a new optimization problem with 30,000 unknowns, whose solution is a scoring function. We trained the scoring function on 6,000 structures of protein-ligand complexes of high accuracy from the PDB database. Despite the very high dimensionality of the optimization problem, we manage to solve it on a desktop computer in just a few hours.

Our scoring function has three major applications in drug-design:

- Docking: determination of the binding site of a ligand bound to a protein.
- Ranking: identifying a set of ligands with the highest binding affinity for the given protein target by screening a large ligand database.
- Binding constants prediction: prediction of the absolute value of the binding constant of a protein-ligand complex.

The success rates of our method rank it among the top three methods currently available. Thus, we believe that our scoring function is the first one that performs well in all three major applications in drug-design.

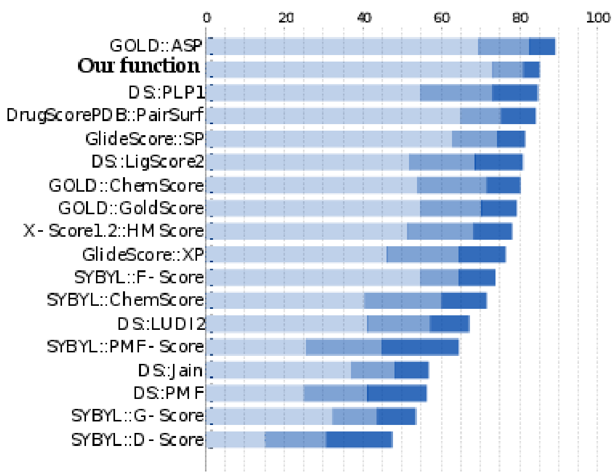


Figure 11. Comparison of the success rates of scoring functions when the best-scored binding pose differs from the true one by RMSD < 1.0 Å (light bars), < 2.0 Å (darker bars) or < 3.0 Å (the darkest bars), respectively. Scoring functions are ranked by success rates when the ligand binding pose is found within RMSD < 3.0 Å.

6.4.2. DockTrina

Participants: Sergei Grudinin, Petr Popov.

We derived analytical formulas for fast evaluation of the Root-Mean-Square-Deviation (RMSD) between rigid protein structures. This work resulted in a RMSD library containing algorithms to calculate the RMSD between two proteins in constant time. Based on this library we introduced an efficient algorithm to predict triangular protein structures and implemented it into the DockTrina software. We collected bound benchmarks of 220 protein trimers with and without symmetry properties from the Protein Data Bank and demonstrated the superiority of DockTrina over standard combinatorial algorithms aimed at predicting nonsymmetrical protein trimers.

6.4.3. Machine Learning for Structural Biology

Participants: Sergei Grudinin, Petr Popov, Mathias Louboutin.

We developed a new formulation of the machine learning optimization problem to predict protein–protein interactions. We implemented several optimization strategies, both in *d*ual and *p*rimal. We studied the effect of different types of loss-functions on the quality of the prediction. We also tested the efficiency of three descent algorithms, Nesterov descent, gradient descent, and stochastic descent. We demonstrated that generally, primal optimization is faster compared to dual optimization. In the primal, Nesterov descent has a better convergence compared to the gradient descent. Finally, stochastic algorithms often provide a better convergence compared to deterministic algorithms. All the studied algorithms were implemented as a stand-alone library.

6.5. Software Engineering

Participants: Jocelyn Gate, Stephane Redon.

We have continued the development of SAMSON, our open-architecture platform for modeling and simulation of nanosystems (SAMSON: Software for Adaptive Modeling and Simulation Of Nanosystems). The interface has been improved:

- The visualization of the data graph has been improved. Users may now drag and drop models and parts between layers, as well as directly drag and drop files into SAMSON.
- The undo/redo stack can now be visualized.
- We have begun to work on selection and highlighting.

The software engineering process has been improved as well, in particular to help base and modules developers:

- We have reorganized the file hierarchy so that modules can have associated data.
- We have developed a system to build SAMSON automatically on virtual machines (e.g., ubuntu 12.04 32bit, ubuntu 12.04 64 bit, fedora 17 32 bit, etc.).
- Tools have been created to let modules developers easily write new modules.
- We have begun to develop a mechanism to make it easy to install and update SAMSON automatically.

We have also developed several *SAMSON apps* to test various concepts, including scripting, manipulating molecules with haptic feedback, etc. Figure 12 shows the current user interface of SAMSON.

We have deposited the first version of SAMSON's code base at the APP ("Agence de Protection des Programmes").

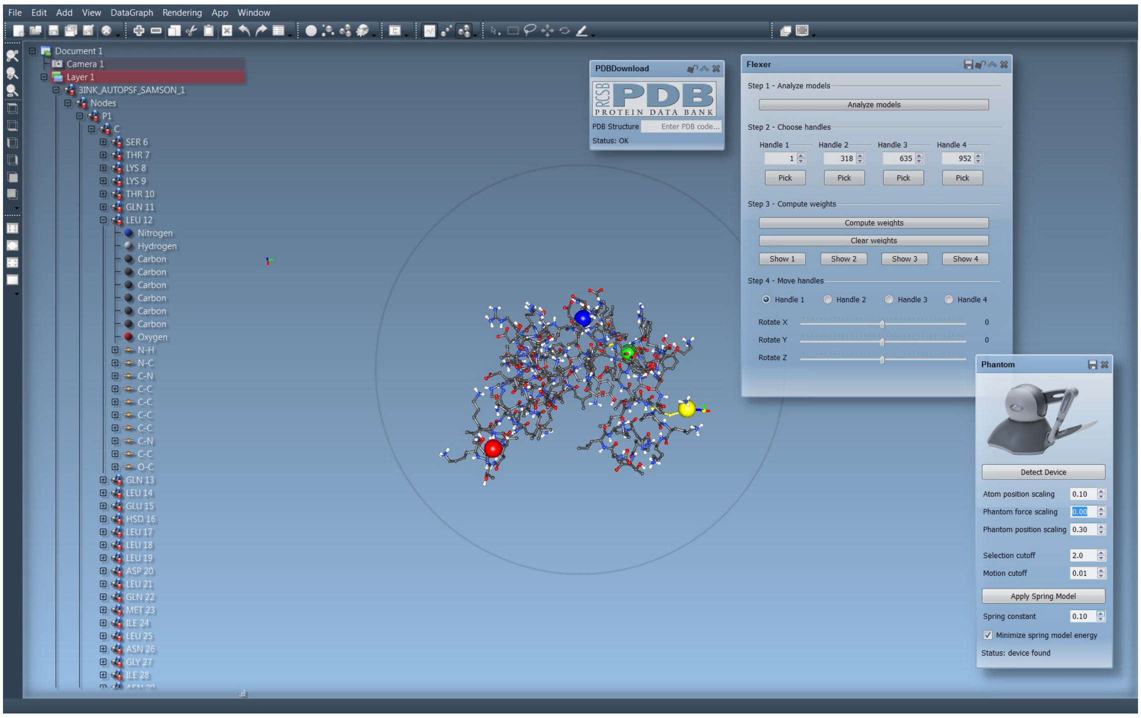


Figure 12. The current user interface of SAMSON, showing an app to download molecules directly from the Protein Data Bank, an app to deform molecules, and an app for haptic interaction. The data graph on the left shows the hierarchical structure of the data graph.

NECS Project-Team

6. New Results

6.1. Communication and control co-design for networked systems

6.1.1. Energy-aware communication and control co-design in wireless networked control systems

Participants: C. Canudas de Wit [Contact person], N. Cardoso de Castro, F. Garin, D. Quevedo [Newcastle Univ., Australia].

This work is the topic of the PhD thesis of N. Cardoso de Castro [12]. We have considered an event-based approach to energy-efficient management of the radio chip in the sensor node of a wireless networked control system.. Indeed, as we had pointed out in the review paper [63], the radio is the main energy consumer, and intermittent data transmission allows one to reduce the use of the radio. While the existing literature in the control community on event-based control only addresses policies using two radio-modes (Transmitting/Sleep), our work follows some considerations on the radio-chip modes well-known in the communication networks literature, and introduces various radio-modes: different 'idle' non-transmitting modes, where only part of the radio-chip is switched off (thus consuming more energy than 'Sleep', but allowing for faster transition to transmission), and various transmitting modes, with different power levels. We propose an event-based radio-mode switching policy, which allows to perform a trade-off between energy saving and performance of the control application. To this end, a switched model describes the system, taking into account control and communication. The optimal switching policy is computed using Dynamic Programming, considering a cost either over an infinite time-horizon [31] or over a finite receding horizon [32].

6.1.2. System-theoretic analysis of modern error correcting codes (serial turbo codes)

Participants: F. Garin [Contact person], G. Como [Lund Univ., Sweden], F. Fagnani [Polit. Torino, Italy].

Serial turbo codes are a family of codes for error correction in point-to-point digital communication. The encoder can be described as the composition of three linear maps, the intermediate one being a permutation (called interleaver) while the inner and outer one are convolutional codes, i.e., linear dynamical systems where state, input and output belong to a vector space over the finite field $GF(2)$. The decoding is performed with iterative low-complexity algorithms which give a good approximation of the optimal maximum-likelihood (ML) decoder. Using system-theoretic properties of the constituent convolutional codes and probabilistic arguments, we study the average and the typical behavior of ensembles of such codes (with fixed convolutional codes, and random interleaver), asymptotically in the block-length [18]. We disprove the common conjecture that the typical behavior concentrates around the average: indeed, the average error decays polynomially in the block-length N, while the typical code has a faster error decay (exponential in some fractional power of N); however, the typical-code analysis confirms the same design parameters for the convolutional codes that were already suggested by the study of the ensemble average: free distance of the outer encoder, and effective free distance of the inner encoder.

6.2. Networked systems and Graph analysis

6.2.1. Observability in consensus networks

Participants: A. Kibangou [Contact person], C. Commault [Gipsa-Lab].

Studying the observability problem of a system consists in answering the question: is it possible, for a given node, to reconstruct the entire network state just from its own measurements and those of its neighbors ?

Studying observability for arbitrary graphs is particularly a tough task. Therefore, studies are generally restricted to some families of graphs. For instance, recently, observability has been studied in [70] for paths and circular graphs where the study was carried out based on rules on number theory. Herein, we have considered families of graphs admitting an association scheme [62] such that strongly regular graphs and distance regular graphs. The regularity properties of these kinds of graphs can particularly be useful for robustifying the network as for cryptographic systems [79]. Based on the so-called Bose-Mesner algebra [60], we have stated observability conditions on consensus networks modeled with graphs modeled with strongly regular graphs and distance regular graphs. For this purpose, we have introduced the notion of local observability bipartite graph that allows characterizing the observability in consensus networks. We have shown that the observability condition is given by the nullity of the so-called local bipartite observability graph. When the nullity of the graph cannot be derived directly from the structure of the local bipartite observability graph, the rank of the associated bi-adjacency matrix allows evaluating the observability; the bi-adjacency matrix of the so-called local bipartite observability graph must be full column rank for guaranteeing observability. From this general necessary and sufficient condition, we have deduced sufficient conditions for strongly regular graphs and distance regular graphs. In particular, we have shown that observability is ensured in such graphs only if $DK \geq N-1$ where D is the number of classes of the association scheme, N the number of nodes, and K the valency of the graph, i.e. the cardinality of the neighborhood.

6.2.2. Distributed graph discovery

Participants: A. Kibangou [Contact person], F. Garin [Contact person], C. Commault [Gipsa-Lab], D. Tran, D. Varagnolo [KTH], K.H. Johansson [KTH].

We have studied the problem of estimating the eigenvalues of the Laplacian matrix associated with a graph modeling the interconnections between the nodes of a given network. Two approaches have been developed. For the first one [38], based on properties of the observability matrix, we have shown that Laplacian eigenvalues can be recovered by solving a local eigenvalue decomposition on an appropriately constructed matrix of observed data. Unlike FFT based methods recently proposed in the literature (see [65], [73]), in our proposed method we are also able to estimate the multiplicities of the eigenvalues. However, this method is only applicable to networks having nodes with sufficient storage and computation capabilities. That's why we have proposed a second method requiring much less computation and storage capabilities in [76]. Based on a recent result showing that the average consensus matrix can be factored in D Laplacian based consensus matrices, where D stands for the number of nonzero distinct Laplacian eigenvalues [40], we have shown how carrying out such a factorization in a fully distributed way. The proposed solution results on a distributed solution of a constrained consensus problem.

The availability of information on the communication topology of a wireless sensor network is essential for the design of the estimation algorithms. In the context of distributed self-organized sensor networks, there is no central unit with the knowledge of the network, and the agents must run some distributed network-discovery algorithms. This is particularly difficult in the case when the agents do not have or do not want to disclose their identifiers (IDs), either for technological reasons (in time-varying self-organized networks, assigning unique identifiers to agents is a challenge) or for privacy concerns. In a recent work [78] the authors proposed an algorithm which allows each agent to find an estimate of the number of agents in the network, in an anonymous way. Such an algorithm is based on the generation of pseudo-random numbers, on some consensus algorithms (for distributed computation either of average or of maximum), and on statistical inference. In our work [37], we show how the same algorithm, with some minor modifications, can provide more information: approximations of each node's eccentricity, of the graph diameter and of the graph radius. We study the quality of such approximations, providing tight bounds on the error.

6.3. Distributed methods for control

6.3.1. Distributed control

Participants: A. Seuret [Contact person], G. Rodrigues de Campos, L. Brinon-Arranz, D.V. Dimarogonas [KTH], K.H. Johansson [KTH].

Another particular effort has been provided to the design of distributed control laws for multi-agents systems. Three main contributions have been produced and can be summarized as follows.

In [44], a new consensus algorithms for heterogeneous multi-agent systems is provided. A control strategy based on a consensus algorithm which is decoupled from the original systems is proposed. Consequently, its major advantage remains in the separation of the stability analysis of each subsystem and the distributed control algorithm. It is shown that our method allows using classical distributed consensus algorithms such as simple integrator consensus (with or without delay) and distributed consensus filter algorithms.

For many multi-robot applications it is interesting to impose a particular configuration for the robotic agents. This paper discusses the design and analysis of a distributed algorithm for the compact deployment of agents, where the behavior of each vehicle is only dependent on local information. The objective of the paper [72] is to achieve the most compact formation possible. To solve this problem we propose, in a first step, two uncorrelated controllers: one designed for dispersion with connectivity maintenance and a second designed to minimize inter-agent angles. An improved controller including variable gains, particularly designed to avoid singular configurations, is also provided. Lastly, we propose a sequential strategy composed of the two previously mentioned controllers and a stability analysis based on hybrid systems theory. Finally, some simulation results for different configurations supporting our theoretical results are presented.

6.3.2. Collaborative source seeking control

Participants: C. Canudas [Contact person], R. Fabbiano, F. Garin.

The problem of source localization consists in finding the point or the spatial region from which a quantity of interest is being emitted; this goal can be pursued by one or several agents possibly cooperating each other. Source-seeking agents can be fixed sensors, that collect and exchange some information about the signal field and try to identify the position of the source (or the smallest region in which it is included), or moving devices equipped with one or more sensors, that physically reach the source in an individual or cooperative way.

Within the FeedNetBack European project, we have addressed the problem of collaborative source seeking with a fleet of autonomous underwater vehicles (UAVs). This topic was explored in the PhD thesis of Lara Brinon [61], where a solution was proposed, based on circular formations with the center of the formation following a 2-dimensional movement in the direction of the gradient of the source. The gradient computation was achieved through an approximation using the point-wise measurements from the various vehicles.

In a more recent work [29], we leave temporarily aside all issues of coordination and communication failures well-addressed in [61], and we focus on the gradient computation formula. Under some assumptions on the source emission (isotropic diffusive source in steady-state, whose solution satisfies the Laplace equation), we show that there is an exact integral formula (based on the Poisson integral of harmonic functions) for the computation of the gradient at the center of a circle, using pointwise measurements along the circumference. This approach has two main advantages: it can be generalized in three (or more) dimensions, and it allows to compute also higher-order derivatives, which allow to find higher-order control laws, useful e.g. for non-holonomic vehicles. A relevant property is that such an integral formula exploits mathematical properties of the source density distribution (the fact that it is harmonic), but does not require the knowledge of an explicit expression for the density function. This makes our approach different from the main source-seeking techniques present in the literature, which either are based on a specific knowledge of the solution of the diffusion process, or make use of an extremum-seeking approach, exciting the system with a periodic signal so as to explore the field and collect enough information to reconstruct the gradient of the quantity of interest.

The latter work is part of the research of Ruggero Fabbiano during his Ph.D. studies.

6.3.3. Distributed real-time Simulation of numerical models

Participants: D. Simon [Contact person], A. Ben Khaled [IFPEN], M. Ben Gaid [IFPEN].

The need of quick innovation in the automotive domain made simulation necessary at early stages of the development cycle. Vehicles and powertrains are complex systems where different domains are involved. Representative phenomenological models of powertrains have been developed and have been used in the design phase under domain dedicated tools. However, their use for controls validation using Model-In-the-Loop (MIL) and Hardware-In-the-Loop (HIL) was prevented due to performance limitation of widely used single-solver/single-core simulation approaches.

Multicore simulation for complex systems has been studied with a focus on simulation duration speedup. The methodology of parallelization across the model has been selected for such problem where strong interactions between the model components are observed. The current study showed that decoupling the model parts by relaxing their data dependencies is promising in term of simulation speed (by increasing the parallelism) and results accuracy. Besides, tests results on engine model showed that, with the model partitioning, it is possible to use efficiently variable-step solvers thanks to the decrease of the number of discontinuities, so the number of integration interrupts, in each subsystem [26].

Further work will investigate in the combination of the use of variable-step solvers in split model with the use of multicore architecture for parallel computing, in order to improve the simulation speedup while keeping results accuracy under control.

6.4. Distributed average consensus algorithms

6.4.1. Finite-time average consensus protocols

Participants: A. Kibangou [Contact person], D. Tran.

Nowadays, several distributed estimation algorithms are based on the average consensus concept. Average consensus can be reached using a linear iterations scheme where each node repeatedly updates its value as a weighted linear combination of its own value and those of its neighbors. The main benefit of using a linear iterations scheme is that, at each time-step, each node only has to transmit a single value to each of its neighbors. Based on such a scheme, several algorithms have been proposed in the literature. However, in the majority of the proposed algorithms the weights are chosen so that all the nodes asymptotically converge to the same value. Sometimes, consensus can be embedded as a step of more sophisticated distributed. Obviously, asymptotic convergence is not suitable for these kinds of distributed methods. Therefore, it is interesting to address the question of exact consensus in finite-time. For time-invariant network topologies and in the perfect information exchange case, i.e. without channel noise nor quantization, we have shown that the finite-time average consensus problem can be solved as a matrix factorization problem with joint diagonalizable matrices depending on the Graph Laplacian eigenvalues [40], [39]. Moreover, the number of iterations is equal to the number of distinct nonzero eigenvalues of the graph Laplacian matrix. The design of such a protocol requires the knowledge of the Laplacian spectrum, which can be carried out in a distributed way (see [65], [73], [76]). In [77], the matrix factorization problem is solved in a distributed way. In particular a learning method was proposed and the optimization problem was solved by means of distributed gradient backpropagation algorithms. Unlike the method in [40], the factor matrices are not necessarily symmetric and the number of these factor matrices is exactly equal to the diameter of the graph.

6.4.2. Quadratic indices for performance evaluation of consensus algorithms

Participants: F. Garin [Contact person], S. Zampieri [Università di Padova], E. Lovisari [Università di Padova and Lund Univ.].

Traditional analysis of linear average-consensus algorithms studies, for a given communication graph, the convergence rate, given by the essential spectral radius of the transition matrix (i.e., the second largest eigenvalues' modulus). For many graph families, such analysis predicts a performance which degrades when the number of agents grows, basically because spreading information across a larger graph requires a longer time. However, when considering other well-known quadratic performance indices (involving all the eigenvalues of the transition matrix), the scaling law with respect to the number of agents can be different. This is consistent with the fact that, in many applications, for example in estimation problems, it is natural to expect that a larger number of cooperating agents has a positive, not a negative effect on performance. It is natural to use a different performance measure when the algorithm is used for different purposes, e.g., within a distributed estimation or control algorithm. Examples of various relevant costs can be found in the book chapter [66] and in the references therein.

We are interested in evaluating the effect of the topology of the communication graph on performance, in particular for large-scale graphs. Motivated by the study of wireless sensor networks, our main objective is to understand the limitations which arise when agents are limited to truly local interactions, i.e., the neighborhoods are determined by being ‘near’ in a geometric (Euclidean) way, differently from graphs with few but possibly ‘distant’ connections, such as in small world models. At first [19] we consider graphs which are regular lattices (infinite lattices, or grids on tori, or grids on hyper-cubes), which are examples of geometrically local interactions, but also have a very rich structure: their symmetries allow to exploit powerful algebraic tools, such as the discrete Fourier transform over rings, to compute their eigenvalues, and then find bounds on the associated costs. Then, we extend the results to a more general class of graphs, thus showing that the behavior of lattices is mainly due to the local nature of interactions and not to the spatial invariance (the richness of the automorphism group). To do so, we exploit the analogy between reversible Markov chains and resistive electrical networks, which allows to study some perturbed grids, with less regularity but still exhibiting the same dimension-dependent asymptotic behavior. This latter work is part of the Ph.D. thesis of E. Lovisari at University of Padova, Italy, and the topic of a journal paper in preparation.

6.5. Distributed Estimation and Data fusion

6.5.1. Distributed joint state and input estimation

Participants: A. Kibangou [Contact person], F. Garin [Contact person], A. Esna Ashari.

Three consensus-based distributed algorithms have been developed for joint state and input estimation in discrete-time systems. The methods are proper substitutes for distributed Kalman filter in the case in which there are additive faults to the system. Previously developed centralized estimation methods have been reformulated so that the estimator can be used for distributed sensor networks. These new forms are similar to the information form of Kalman filter [34], [35]. The new forms can be used to propose distributed algorithms based on the consensus of the nodes on calculation of some matrices and vectors. Also a second algorithm is proposed, based on the consensus of the local estimators on local state estimations. This algorithm has less computation effort than the first, but gives a sub-optimal solution in the sense of covariance error. Finally, a third method based on covariance intersection method for diffusing local estimations was proposed in addition. This method also provides a sub-optimal solution. Compared with the second approach, the diffusion of local data is less complicated, however it requires more message communication between nodes.

6.5.2. Data fusion approaches for motion Capture by Inertial and Magnetic Sensors

Participants: H. Fourati [Contact person], A. Makni.

We are interested to motion capture (or attitude) by fusing Inertial and Magnetic Sensors. In [17], we present a viable quaternion-based Complementary Observer (CO) which is designed for rigid body attitude estimation. We claim that this approach is an alternative one to overcome the limitations of the Extended Kalman Filter (EKF). The CO processes data from a small inertial/magnetic sensor module containing tri-axial angular rate sensors, accelerometers, and magnetometers, without resorting to GPS data. The proposed algorithm incorporates a motion kinematic model and adopts a two-layer filter architecture. In the latter, the Levenberg Marquardt Algorithm (LMA) pre-processes acceleration and local magnetic field measurements, to produce what will be called the system’s output. The system’s output together with the angular rate measurements will become measurement signals for the CO. In this way, the overall CO design is greatly simplified. The efficiency of the CO is experimentally investigated through an industrial robot and a commercial IMU during human segment motion exercises. These results are promising for human motion applications, in particular future ambulatory monitoring. The estimated attitude is used to reconstitute the linear acceleration, linear velocity and finally the 3D position from a usual integration procedure (in the case of foot motion) [36]. The problem of attitude estimation is also recently studied within the PhD thesis of Aida Makni. Our goal is to develop a new attitude estimation methods in the case of aerial vehicles (hexa-rotors) by the use of intermittent measures of gyroscopes with the goal to reduce the energy consumption and to gain in the autonomy of the battery.

6.6. Stability and control design of asynchronous interconnected systems

6.6.1. New approaches for stability analysis of time-delay systems

Participants: A. Seuret [Contact person], F. Gouaisbaut.

A particular attention has been paid to the stability analysis of time delay systems. Indeed delays represent a classical phenomenon which appears in networked control systems cite. This corresponds to the fact that data are not transmitted instantaneously from one node to its neighbors. In this context some effort has been provided in order to reduce the conservatism of the stability conditions. This works represents some fundamental researches to develop accurate stability conditions to networked control systems. More especially we produced a paper [45] which addresses the stability problem of linear time delay system. In the literature, the most popular approach to tackle this problem relies on the use of Lyapunov-Krasovskii functionals. Many results have proposed new functionals and techniques for deriving less and less conservative stability conditions. Nevertheless, all these approaches use the same trick, the well-known Jensen's inequality which generally induces some conservatism difficult to overcome. In light of those observations, we propose to reduce the conservatism of Lyapunov-Krasovskii functionals by introducing new classes of integral inequalities called Wirtinger's inequalities. This integral type inequality is firstly shown to encompass Jensen's inequality and is then employed to derive new stability conditions. To this end, a slightly modified Lyapunov functional is proposed. Several examples illustrate the effectiveness of our methodology. Further efforts on this topics have been provided and several improved articles are now submitted to servals journals.

6.6.2. Stability and control of asynchronous sampled-data systems

Participants: A. Seuret [Contact person], C. Briat [ETHZ], J. Gomes Da Silva Jr. [UFRGS], M. M. Peet [Illinois Institute of Technology].

Sampled-data systems have been extensively studied in the literature and the references therein. It is now reasonable to design controllers which guarantee the robustness of the solutions of a closed-loop system under periodic samplings. However the case of asynchronous samplings still leads to several open problems. This corresponds to the realistic situation where the difference between two successive sampling instants is time-varying. Several articles drive the problem of time-varying periods based on a discrete-time approach, input delay approach using the framework of Lyapunov-Krasovskii theorem, using the small gain theorem or the analysis of impulsive systems. These last approaches are very relevant to this problem because they cope with time-varying sampling periods as well as with uncertain systems in a simple manner. Nevertheless, these sufficient conditions are still more conservative than discrete-time approaches. In [24], we proposed a novel approach to assess stability of continuous linear systems with sampled-data inputs. The method, which is based on a particular type of functionals, called 'looped-functionals' provided easy tractable stability conditions for the continuous-time model. This method has been extended to various cases dealing with sampled-data systems. Indeed a method to constructs such class of functionals using the Sum of Squares framework was developed in [23]. Another extensions was also proposed in order to include saturations in the actuators [21]. Extensions to the case of communication delays and asynchronous samplings was also provided in [22].

Based on this new type of Lyapunov functional, several works have been provided in the more general context of hybrid system. Indeed sampled-data systems can be seen as a particular type of hybrid systems. This has been provided in several study done by A.R. Teel, Dragan Nesic and many other researchers. Thus the idea was to show that the previous approach was also able to provide efficient stability conditions for impulsive systems [16], [27], [28], [54] or switched systems [53].

6.6.3. Event-based control

Participants: A. Seuret [Contact person], N. Marchand [Gipsa-Lab], C. Prieur [GIPSA-Lab], S. Durand [CINVESTAV].

Usually feedback laws are implemented in a periodic fashion on digital hardware. The main reason for using this periodicity in the hardware comes from the difficulties to analyze the stability of aperiodic or asynchronous systems. However it also seems natural to hold the same control input longer if the system behaves in a suitable way or shorter if the system requires an updated input. In [9], an algorithm is suggested to sample the control input based on the behavior of a Lyapunov-like function. This algorithm is called event-triggered since the Lyapunov-like function directly depends on the state of the systems. Using a Lyapunov-like function, two algorithms for the design of event-triggered algorithm are designed. It is assumed that a stabilizing controller for the continuous control system is given. Both event-triggered algorithms need to consider a closed-loop system with a mixed discrete/continuous dynamics (namely this is a hybrid system). Some numerical simulations illustrate the stability properties of both algorithms. In a future work, the performance issue should be analyzed. It is remarked that the event-triggered algorithms have a different performance. The first one seems to ensure a good speed of convergence on numerical simulations, whereas the second event-triggered algorithm allows less jumps and thus needs to compute less often the control variables. The advantages and disadvantages of each algorithm will be studied more precisely in a future work, for a theoretical point of view (e.g. by estimating a priori the number of switches), or on applications (to understand which algorithm is better depending on the application). Regarding this remark a journal paper has been submitted to IMA Journal of Mathematic Control and Information lately in 2012.

6.6.4. Feedback under slacken real-time

Participants: D. Simon [Contact person], A. Seuret, P. Andrianiaina [AIRBUS].

Robustness in control usually deals with the plant's parameter uncertainties, but the insensitivity or adaptability w.r.t. timing deviations from the theoretical pattern, such as jitter or deadlines misses can be exploited. The interesting point is that a feedback control system which is robust w.r.t. the plants parameters uncertainties is also robust, to some extent, w.r.t. timing deviations. Hence a feedback control system is not as hard as it is often considered in the literature, but should be better considered as *weakly hard*, i.e. able to tolerate specified timing deviations without leaving its requested performance [46].

A weakened implementation scheme for real-time feedback controllers is proposed to reduce the conservatism due to traditional worst-cases considerations. To save wasted computing resources, new real-time scheduling scenarios allowed for reducing the time slots allocated to control tasks below the value of the Worst Case Execution Time which is traditionally used to implement embedded control software. The stability of the control system under occasional deadlines miss is assessed using robustness arguments, using Lyapunov-Krasovskii functionals and LMIs solving based on [10]. The methodology has been successfully assessed for a fighter aircraft pitch controller, which show that the stability of the plant can be kept (and even improved) using the new scheduling schemes using less computing resources than traditional implementations [25], [11].

6.6.5. Varying sampling for LPV systems

Participants: D. Simon [Contact person], O. Sename, E. Roche.

In the context of network-controlled systems the idea of using varying control intervals naturally arises when the available computing power devoted to feedback control is limited, e.g. in embedded systems. It can be easily shown that decreasing the control frequency directly decreases the amount of computing needed for control. However, the stability of the feedback controller under varying sampling must be assessed for all the allowed variations of the sampling intervals [8].

The Linear Fractional Transform (LFR) formulation is widely used in robust analysis to study the influence of the plant's uncertain parameters on the stability and performances of a closed-loop system. Usually it is used to build a parameter dependent model of a dynamical system, depending on a known set of parameters. Here the set of varying parameters has been extended with the sampling interval of the control system, thus allowing to handle both varying sampling and plants uncertainties in a single framework (Figure 6).

Here, P_d is a on-line discretized model of the plant, Δ represents the uncertain parameters of the plant and δ is the variation of the sampling interval around its nominal value. From this model a robust controller can be synthesized, enforcing the control system stability for all variations of the sampling interval inside a predefined range.

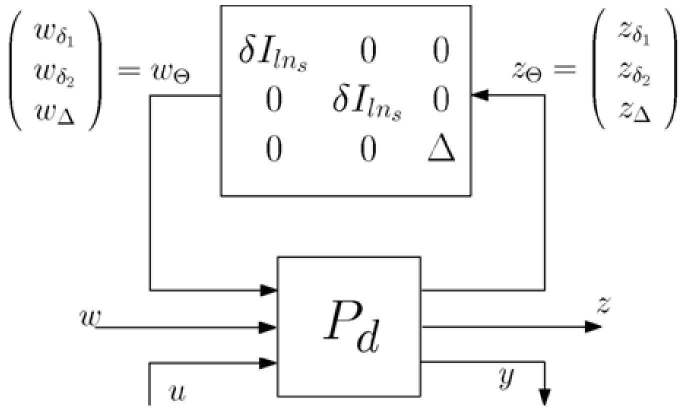


Figure 6. LFR system depending on system parameters and sampling interval variations

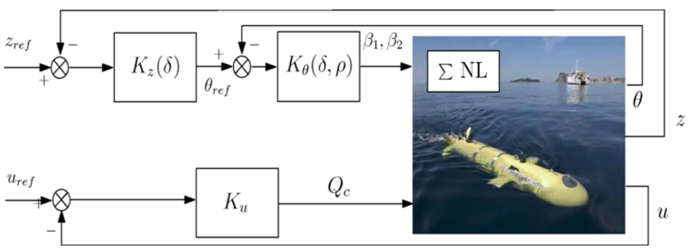


Figure 7. Control of an AUV

The approach have been successfully applied to the pitch and altitude control of a non-linear autonomous underwater vehicle, where the source of sampling variations comes from the altitude ultrasonics sensors [43]. However the approach still suffers from conservatism for which improvements using full block multipliers, or parameter-dependent Lyapunov functions, have been investigated [58].

6.7. Vehicular transportation systems

6.7.1. Traffic estimation and prediction

Participants: C. Canudas de Wit [Contact person], A. Kibangou, L. Leon Ojeda, F. Morbidi.

Reconstructing densities in portions of the road links not equipped with sensors constitutes an important task in traffic estimation, forecasting, and control problems. Among many other approaches, model-based observers is one popular technique to build this information. They can also be understood as *virtual sensors* deployed inside of the cells not equipped with *true sensors*. They are used to better track, in real-time, density variations with a fine degree of granularity in the space, as the *virtual cells* can be selected as small as desired. In [30], a graph constrained-CTM observer was introduced. It allows reconstructing rather accurately the internal states (densities) of a road portion not equipped with sensors. This strategy for real-time density estimation was applied on Grenoble South Ring. Simulation results exhibit that the measured densities obtained from the traffic simulator Aimsun and the estimated densities agree closely. In [69], this observer has been associated with an adaptive Kalman filtering approach for traffic prediction in terms of travel time. The adaptive Kalman filtering approach was also been used for predicting input flows in [68].

6.7.2. Traffic control

Participants: C. Canudas de Wit [Contact person], D. Pisarski.

The problem of equilibrium points for the Cell Transmission Model was studied in [42]. The structure of equilibrium sets was analyzed in terms of model parameters and boundary conditions. The goal was to determine constant input flows, so that the resultant steady state of vehicle density was uniformly distributed along a freeway. The necessary and sufficient conditions for the existence of one-to-one relation between input flow and density were derived. The equilibrium sets were described by formulas that allow to design a desired balanced density. A numerical example for the case of a two-cell system was presented. In [41], the problem of optimal balancing of traffic density distributions was explored. The optimization was carried out over the sets of equilibrium points for the Cell Transmission Traffic Model. The goal was to find the optimal balanced density distribution, that maximizes both the Total Travel Distance and the total input flow. The optimization was executed in two steps. At the first step, a nonlinear problem to find a uniform density distribution that maximizes the Total Travel Distance was solved . The second step was to solve a quadratic problem reflecting the trade-off between density balance and input flow maximization. At both steps, decomposition methods were used. The computational algorithms associated to such a problem were given. Finally, in [71], the application of the idea of optimal balancing of traffic density distribution was presented. It was implemented to the Grenoble South Ring in the context of the Grenoble Traffic Lab. The traffic on the ring is represented by the Cell Transmission Model that was tuned by using real data and Aimsun micro-simulator. A special attention was paid to the calibration of a flow merging model. A large-scale optimization problem was solved by using advanced combinatorial procedures. The main difficulties in the implementation as well as the limitations of the designed software were highlighted. Finally, the results of different traffic scenarios on the Grenoble South Ring were presented.

6.7.3. Vehicle control for disabled people

Participants: C. Canudas de Wit [Contact person], V. Ciarla, J. Dumon, F. Quaine [UJF], V. Cahouet [UJF].

The typical architecture of an Electric Power Assistance Steering (EPAS) system includes a static map to provide the correct amplification to the drivers exerted torque. In literature, it is generally known as booster curve. This work concerns the study of the amplification criteria, that are commonly used to these booster curves. The basic concepts of the Electric Power Steering (EPS) systems with a realistic model for the friction contact, that acts on the wheels are discussed. A relation between the assistance and the driver's torque is provided, under the hypothesis of a position-oriented control of the movement and the Stevens' power law [33]. In current works, we want to modify the general architecture of the EPAS system for people driving with two arms. For this purpose, we insert two additional blocks: the first one provides an estimation of the gravitational torque due to the weight of the driver's arm while the second gets as inputs the total driver's torque and the estimated gravitational torque in order to update the driver's torque with the gravitational torque. The updated measure is then given as input to the booster curve for deriving the correct assistance.

6.7.4. Control of communicating vehicles in urban environment

Participants: C. Canudas de Wit [Contact person], G. de Nunzio.

For a given vehicle there are different ways to travel on a given distance in a given time, corresponding to different levels of energy consumption; therefore, there is an energy-optimal trajectory. Advising the driver via a suitable interface can reduce the energy consumed during the travel, and thus improve the energy efficiency: this is the principle of eco driving. In urban areas, the optimal trajectory of the vehicle depends on interactions with other vehicles, but also on passive signs (panels, priorities, etc.) and active signs (traffic lights); in each case, constraints are imposed on the command (vehicle speed). From the infrastructure perspective, traffic control in urban areas consists in determining the state of traffic signals in order to solve an optimization problem, for example minimizing travel time of vehicles in the road network. If all the vehicles can communicate with one another and with the active infrastructure (traffic lights), we can imagine benefits for each of the two problems which can be considered as a whole: on the one hand, for vehicles, more information is available that can be integrated into the online optimization problem; on the other hand, there are new measures and new commands available to control traffic. Indeed, the estimation of the traffic is no longer necessary, as the position and speed of approaching vehicles is known. More importantly, the traffic manager can send instructions to the vehicle. The aim of the research is to evaluate the potential in terms of energy saving and traffic improvement made possible by communicating vehicle. This work is carried out in collaboration with IFP in the framework of a CIFRE thesis.

NON-A Project-Team

5. New Results

5.1. Model-free control

Participants: Cédric Join, Samer Riachy.

The achievements obtained in 2012 are as follows:

- The model-free control approach is applied to a complex nonlinear model describing the dynamics of a traffic flow in [24]. The robustness with respect to external disturbances is shown by numerical simulations.
- Model-free control is applied to a magnetic bearing in [56], which is a quite important industrial device. The experimental results are compared to those obtained via other control techniques.
- ”Model-free” control and the related ”intelligent” proportional-integral controllers are successfully applied to freeway ramp metering control in [47]. Implementing the proposed control strategy is straightforward. Numerical simulations need the identification of quite complex quantities like the free flow speed and the critical density. This is achieved due to new estimation techniques, where the differentiation of noisy signals plays a key role.

5.2. Algebraic technique for estimation, differentiation and its applications

Participants: Cédric Join, Mamadou Mboup, Wilfrid Perruquetti, Rosane Ushirobira, Olivier Gibaru.

Elementary techniques from operational calculus, differential algebra, and noncommutative algebra lead to a new algebraic approach for estimation and detection. It is investigated in various areas of applied sciences and engineering. The following lists only some applications:

- The paper [30] proposes an algebraic method to fault diagnosis for uncertain linear systems. The main advantage of this new approach is to realize fault diagnosis only from knowledge of input and output measurements without identifying explicitly model parameters. Using tools and results of algebraic identification and pseudospectra analysis, the issues of robustness of the proposed approach compared to the model order and noise measurement are examined.
- The aim of [79], [84] is to develop an algebraic approach to estimate human posture in the sagittal plane using inertial measurement unit providing accelerations and angular velocities. For this purpose the issue of the estimation of the amplitude, frequency and phase is addressed for a biased and noisy sum of three sinusoidal waveform signals on a moving time horizon. Since the length of the time window is small, the estimation must be done within a fraction of the signal’s period. The problem is solved via algebraic techniques.
- An application of algebraic estimation approach for estimation of option pricing and dynamic hedging is given in [66].
- A model-based online fault-diagnosis scheme for an electromagnetically supported plate is presented in [73] as an example of a nonlinear and open-loop unstable system. First, residuals for sensor as well as for actuator faults are generated using algebraic derivative estimators. Then, the robust detection and isolation of step-like sensor and actuator faults is presented.
- The paper [57] uses the extreme value theory for threshold selection in a previously proposed algebraic spike detection method. The algebraic method characterizes the occurrence of a spike by an irregularity in the neural signal and devises a nonlinear (Volterra) filter which enhances the presence of such irregularities.

- The papers [39], [40] generalize the algebraic method from the integer order to the fractional order for estimating the fractional order derivatives of noisy signals. The proposed fractional order differentiator is deduced from the Jacobi orthogonal polynomial filter and the Riemann-Liouville fractional order derivative definition. Exact and simple formula for this differentiator is given where an integral formula involving Jacobi polynomials and the noisy signal is used without complex mathematical deduction. Hence, it can be used both for continuous-time and discrete-time models. The comparison between our differentiator and the recently introduced digital fractional order Savitzky-Golay differentiator is given in numerical simulations so as to show its accuracy and robustness with respect to corrupting noises.

5.3. Observability and observer design for nonlinear systems

Participants: Jean-Pierre Barbot, Wilfrid Perruquetti, Gang Zheng, Denis Efimov.

Observability analysis and observer design are important issues in the field of control theory. Some recent results are listed below:

- The problem of observer design for fault detection in a class of nonlinear systems subject to parametric and signal uncertainties is studied in [22]. The design procedure includes formalized optimization of observer free parameters in terms of trade-offs for fault detection performance and robustness to external disturbances and model uncertainties. The technique makes use of some monotonicity conditions imposed on the estimation error dynamics. Efficiency of the proposed approach is demonstrated through the Oscillatory Failure Case in aircraft control surface servo-loops.
- An algorithm for the frequency and bias identification of a harmonic signal is presented in [14], [15]. The solution is based on an adaptive observer technique and the hybrid systems method.
- An influence of a singular manifold of non observable states on reconstruction of chaotic attractors is analysed in [25]. The probability of visits of the observability singularity manifold and the relative time spent in the observability singularity manifold are introduced.
- In [36], the cluster structured sparse signals are investigated. Under the framework of Bayesian compressive sensing, a hierarchical Bayesian model is employed to model both the sparse prior and cluster prior, then Markov Chain Monte Carlo (MCMC) sampling is implemented for the inference. Unlike the state-of-the-art algorithms, which are also taking into account the cluster prior, the proposed in [36] algorithm solves the inverse problem automatically–prior information on the number of clusters and the size of each cluster is unknown.
- The papers [54], [86], [87] present a new approach for observer design for a class of nonlinear singular systems which can be transformed into a special normal form. The interest of the proposed form is to facilitate the observer synthesis for the studied nonlinear singular systems. Necessary and sufficient geometrical conditions are deduced in order to guarantee the existence of a diffeomorphism, which transforms the studied nonlinear singular systems into the proposed normal form.
- The paper [85] investigates the observer design problem of for linear switched system with disturbance jumps. Detection of active sub-system and finite time estimation of states are respectively discussed. A switched finite time observer is proposed to guarantee the finite time convergence independent of the disturbance jumps.
- The paper [71], [72] proposes a new observer scheme for chaotic and hyperchaotic systems. Firstly, a classical impulsive observer is investigated for Lorenz chaotic system. This approach is based on sufficient conditions for stability of impulsive dynamical systems. After, an hybrid observer is proposed for hypoerchaotic systems. In the paper [70], a new method of strange attractor identification, under sparse measurement, is proposed this method is based on the concept of compressive sensing. For this, some particular impulsive observers have been presented with a decision scheme linked to diagnosis method, the identification of the strange attractor and state observation are done.

- The problem of state reconstruction for nonlinear differential-algebraic systems with unknown inputs is studied in [51].
- In the paper [26] the design of observers for nonlinear systems with unknown, time-varying, bounded delays, on both state and input for a class of nonlinear systems is proposed. Furthermore, the feasibility of the proposed strategy is illustrated by a numerical example.

5.4. Sliding mode control estimation

Participants: Jean-Pierre Barbot, Wilfrid Perruquetti, Denis Efimov, Thierry Floquet.

Sliding mode algorithms are very popular for finite-time estimation and regulation. The recent results obtained by the group are as follows:

- The issues of a higher order sliding mode controller realization under actuator saturation and quantization have been analysed in [37]. The zig-zag solutions are introduced and analysed.
- The problem of design of interval observers for linear-parameter-varying systems, containing non detectable or non strongly observable parts, is addressed in [18], [63], [62] applying the higher order sliding mode algorithms. Application of sliding mode observers leads to accuracy improvement in the system.
- In [32] an anomaly signal detection in communication networks is studied by control theory techniques. Several classes of sliding mode observers are proposed for a fluid flow model of the transmission control protocol (TCP)/internet protocol network. Comparative simulations via network simulator NS-2 show the enhancement brought by a higher order sliding mode observer. The efficiency of this observer opens the way toward observing traffics with real TCP flow characteristics.
- In [80], [42], [41] the problem of continuous and discrete state estimation for a class of linear switched systems is studied. The class of systems under study can contain non-minimum phase zeros in some of their "operating modes". The conditions for exact reconstruction of the discrete state are given using structural properties of the switched system. The state-space is decomposed into the strongly observable part, the nonstrongly observable part and the unobservable part, to analyze the effect of the unknown inputs. A state observer based on high-order sliding-mode and Luenberger-like observers is proposed. For the case when the exact reconstruction of the state cannot be achieved, the ultimate bounds on the estimation errors are provided. In [41] this technique has been applied to fault detection in switched systems.
- The paper [55] aims, firstly to highlight the possibility of recovering a message included in a chaotic continuous time delay system, secondly to show that it is possible to use the third order sliding mode in order to recover directly all the states and the unknown input (message), thirdly to illustrate the robustness of the proposed observer with respect to a noisy signal. This work is based on the concept of left invertibility and recent advances in sliding mode observers.
- The problem of estimation of discrete and continuous states for switched systems applying higher order sliding mode observers and projection is investigated in the papers [68], [67].

5.5. Non-linear, Sampled and Time-delay systems

Participants: Jean-Pierre Richard, Lotfi Belkoura, Gang Zheng, Denis Efimov, Wilfrid Perruquetti.

Nonlinearities, sampling, quantization and time-delays cause serious obstructions for control and observer design in many fields of techniques and engineering (e.g. networked and internet systems, distributed systems etc.). The proposed by the team algebraic approach suits well for estimation and regulation in such a type of systems. The recent results are listed below:

- A new type of stability is introduced and its equivalent Lyapunov characterization is presented in [16]. The problem of global stability for the compact set composed of all invariant solutions of a nonlinear system (several equilibriums, for instance) is studied. It is shown that several well-known multi-stable systems satisfy this new stability property.

- A new state-dependent sampling control is proposed in [23], [65], which enlarges the sampling intervals of state feedback control. The case of linear time invariant systems with time delays is considered that guarantees the exponential stability of the system origin for a chosen decay rate. The approach is based on LMIs obtained from the sufficient Lyapunov-Razumikhin stability conditions.
- Nonlinear feedback design for fixed-time stabilization of linear control systems is studied in [31]. Nonlinear control algorithms of two types are presented for uncertain linear plants. Controllers of the first type are stabilizing polynomial feedbacks that allow to adjust a guaranteed convergence time of system trajectories into selected neighborhood of the origin independently on initial conditions. Controllers of the second type are modifications of the second order sliding mode control algorithms. They provide global finite-time stability of the closed-loop system and allow to adjust a guaranteed settling time independently on initial conditions. Control algorithms are presented for both single-input and multi-input systems.
- The problem of natural wave control is addressed in [17], which involves steering a lattice of oscillators towards a desired natural (i.e. zero-input) assignment of energy and phase across the lattice. This problem is formulated and solved for lattices of linear oscillators via a passivity-based approach.
- The verification problems for transition systems enriched with a metric structure is analysed in [27]. The main novelty compared to an algorithm presented recently by Lerda et al. [2008] consists in introducing a tuning parameter, which improves the performance drastically. A procedure that allows one to prove unbounded safety from the result of the bounded safety algorithm via a refinement step is also established. The algorithm to handle bounded liveness verification is adapted.
- The problem of finite-time output stabilization of the double integrator is addressed in [52] applying the homogeneity approach. A homogeneous controller and a homogeneous observer are designed (for different degree of homogeneity) ensuring the finite-time stabilization. Their combination under mild conditions is shown to stay homogeneous and finite-time stable as well.
- The notes [76], [77] are dedicated to the stability analysis of bilinear sampled-data systems, controlled via a linear state feedback static controller. A zero order hold device is used. The purpose is to find a constructive way to calculate the maximum allowable sampling period (MASP) that guarantees the local stability of the system. The proposed stability conditions are formulated as linear matrix inequalities (LMI).
- The works [75], [74] concern the adaptation of sampling times for linear time invariant systems controlled by state feedback. Complementary to various works that guarantee stabilization independently of changes in the sampling rate, there the conditions to design stabilizing sequences of sampling instants is provided. In order to reduce the number of these sampling instants, a dynamic scheduling algorithm optimizes, over a given sampling horizon, a sampling sequence depending on the system state value. The proofs are inspired on switched system techniques combining Lyapunov functions and LMI optimization.
- The mechanism of entrainment to natural oscillations in a class of (bio)mechanical systems described by linear models is investigated in [61]. A nonlinear control strategy (based on the speed gradient control algorithm) is analyzed providing the system oscillation in resonance mode with a natural frequency. It ensures an energy-optimal entrainment performance robustly against perturbations in system parameters in a finite time.
- The paper [29] considers a networked control loop, where the plant is a "slave" part, and the remote controller and observer constitute the "master". Since the performance of Networked Control Systems (NCS) depends on the Quality of Service (QoS) available from the network, a controller is designed that takes into account qualitative information on the QoS in real time.
- In the paper [50], the theory of non-commutative rings allows determining whether or not there exists an equation called algebraically essential in order to estimate the delay on a nonlinear system. From this equation, it is shown that this equation is generally not enough to guarantee the delay estimation, thus the notion of persistent signal with respect to delay estimation is introduced.

Furthermore, based on the definitions of algebraically essential equation and of persistent signal, a delay estimation algorithm is proposed. Some simulation results have been presented in order to highlight the robustness (with respect to measurement noise) of the proposed algorithm.
- The problem of algebraic identifiability for linear and nonlinear dynamical systems is considered in [88].

5.6. Interval control and estimation

Participants: Denis Efimov, Wilfrid Perruquetti.

In many cases due to parametric and/or signal uncertainties presented in a plant model it is not possible to design a conventional observer, which provides a point-wise estimate of state in a finite time or asymptotically. In this case it is still frequently possible to apply the interval observer techniques, which generate an estimate on the interval of the admissible values of the state at the current instant of time. The recent results are listed below:

- The problem of output stabilization of a class of nonlinear systems subject to parametric and signal uncertainties is studied in [20], [21]. First, an interval observer is designed estimating the set of admissible values for the state. Next, it is proposed to design a control algorithm for the interval observer providing convergence of interval variables to zero, that implies a similar convergence of the state for the original nonlinear system. An application of the proposed technique shows that a robust stabilization can be performed for linear time-varying and linear-parameter-varying systems without assumption that the vector of scheduling parameters is available for measurements.
- The problem of interval observer design for a class of observable nonlinear systems is studied in [33]. It is shown that under some mild conditions a Hurwitz matrix can be transformed to a Hurwitz and Metzler one using a real similarity transformation.
- The work [64] is devoted to interval observer design for Linear Time Varying (LTV) systems and a class of nonlinear time-varying systems in the output canonical form. An interval observer design is feasible if it is possible to calculate the observer gains making the estimation error dynamics cooperative and stable. It has been shown that under some mild conditions the cooperativity of an LTV system can be ensured by a static linear transformation of coordinates. The case of a time-varying transformation for periodic systems is considered in the work [64].
- The problem of actuator fault detection for flat systems using the sliding-mode differentiation and the interval constraint satisfaction technique has been analysed in [43].

5.7. Applications

Participants: Jean-Pierre Richard, Jean-Pierre Barbot, Mamadou Mboup, Gang Zheng, Denis Efimov, Wilfrid Perruquetti, Olivier Gibaru, Samer Riachy.

As it was mentioned, Non-A is a kind of "method-driven" project, which deals with different aspects of finite-time estimation and control. Thus different applications are possible, ones touched this year are as follows:

- The global stabilization of a ball & beam through saturated control, which imposes restrictions on the reactivity of the closed loop, is studied in [91], [81]. A modified design for the classical ball & beam system is presented. The beam is driven by two actuators. In comparison to the classical system, this design offers an additional degree of freedom, which is the vertical motion of the beam. We show that the new design offers the possibility to get rid of the closed loop low reactivity restriction. Two nonlinear controllers to steer the trajectories of the system towards a final desired position are proposed.
- In papers [48], [49] a new class of power converters is studied (Parallel Multicell Chopper). The topology of these chopper is based on a combination of n switching cells interconnected via independent inductors. This type of choppers is a new DC/DC static power converter which has an output current equals to n times the source current where n is the number of cells . After recalling the dynamical equations of the converter, its hybrid dynamical behavior and properties are highlighted. This particular hybrid system induces new and difficult control problems, such problem can be tackled by a new control concept based on Petri net.

- The paper [69] addresses the problem of power management of a hydrogen fuel cell system combined with super capacitors under high load variations in an electric vehicle. The singular perturbation theory is used for the control and coordination of two converters. The Lyapunov theory is used for analysis.
- Combined feedforward/feedback control algorithm for highly nonlinear systems was proposed on the basis of the approximating hybrid model in [28]. The designed MIMO controller enables simultaneous control of the air-to-fuel ratio and torque for injector automobile engines. The theoretical results were validated experimentally with physical cars.
- A spike sorting method for multi-channel recordings is proposed in [35]. The proposed method uses an iterative application of Independent Component Analysis (ICA) algorithm and a deflation technique in two nested loops. The results suggest that the proposed solution significantly improves the performance of ICA in spike sorting.
- In the paper [83] an algorithm for a particular change-point detection problem is proposed, where the frequency band of the signal changes at some points in the time axis. Apart from detecting the change-points, the proposed algorithm is also able to estimate the frequency bands. The main idea of the algorithm is to consider a simple local bandlimited model to represent the input signal in each sliding time window.
- The papers [60], [58], [59] present a new sensorless parameter identification method for permanent magnet stepper motors. Current sensors are assumed available, but position and velocity sensors are not. Data is obtained with open-loop voltage commands at multiple speeds. A new reference frame is proposed that presents advantages similar to the standard $d-q$ frame, but without the need for a position sensor. The method exploits carefully derived linear parameterizations and a least-squares algorithm. In one case, overparameterization is resolved using elimination theory. Overall, the parameters identified using the new procedure are found to be very close to those obtained with position sensors. The approach is potentially applicable to other types of synchronous motors as well.
- In the paper [78], an improvement of the dynamic accuracy of a flexible robot joint is addressed. Based on the observation of the measured axis deformation, a simplified elastic joint model is deduced. In the first step, the non-linear model component's is analyzed and identified in the cases of the gravity bias and the friction term. In the second step, a non asymptotically algebraic fast identification of the oscillatory behavior of the robot axis is introduced. Finally, the performances of the identification approach are exploited in order to improve the dynamic accuracy of a flexible robot axis. This is done experimentally by the combination of the adaptation of the jerk time profile to reduce the end-point vibration and the model-based precompensation of the end-point tracking error.
- Localizability of unicycle mobiles robots is analysed in [82] from an algebraic point of view. A sensibility study leads to a new fusion algorithm in the multi landmark case using as a basis the posture differentiation based estimator.
- The problem of early detection of oscillatory failures for aircrafts is addressed in [38]. The proposed solution is based on a finite-time sliding-mode differentiator and a hybrid optimization scheme.
- The H_∞ control design under time-varying delays and uncertainties, which ensures the stability and performance (synchronization/transparency) between the master and slave manipulators, is proposed in [46], [44], [45]. The design of the controller based on a proposed control scheme, which is performed by using LMI optimization based on Lyapunov-Krasovskii functionals and H_∞ control theory.

OPALE Project-Team

6. New Results

6.1. Mathematical analysis and control of macroscopic traffic flow models

6.1.1. Vehicular traffic

Participants: Maria Laura Delle Monache, Paola Goatin, Mauro Garavello [Piedmont University, Italy], Alexandre Bayen [UC Berkeley, CA, USA].

The activity in traffic flow modeling has being reinforced by the creation of the Associated Team ORESTE between OPALE and the UC Berkeley teams Mobile Millennium and Integrated Corridor Management (ICM) lead by Prof. A. Bayen (see http://www-sop.inria.fr/members/Paola.Goatin/ORESTE/index.html). In this framework, three PhD students from US visited Inria during August and September, and M.L. Delle Monache spent two and half months at UC Berkeley.

During this first year of common research we proposed a new junction model for ramp-metering in the continuous and discrete settings. We focused on a junction consisting in a mainline, an on-ramp and an off-ramp. In particular, we introduced a coupled PDE-ODE model, in which the PDE describes the evolution of the cars flow on the mainline and the ODE describes the evolution of the queue length on the on-ramp, modeled by a buffer, which ensures that boundary conditions are satisfied in strong sense. At the junction we imposed the maximization of the outgoing flux together with a fixed priority parameter for incoming roads. We were able to prove existence and uniqueness of the solution of the corresponding Riemann problem. This approach has then been extended to networks and discretized using the Godunov scheme. The corresponding discrete optimization problem has been solved using the Adjoint Method and it is now being implemented into a MATLAB code. This model will serve as starting point for a subsequent model for optimal rerouting, which includes multi-commodity flow and partial control.

Besides that, we studied a a coupled PDE-ODE system modeling the interaction of a large slow moving vehicle with the surrounding traffic flow. The model consists in a scalar conservation law with moving density constraint describing traffic evolution coupled with an ODE for the slow vehicle trajectory. The constraint location moves due to the surrounding traffic conditions, which in turn are affected by the presence of the slower vehicle, thus resulting in a strong non-trivial coupling. The existence result is given in [60].

The paper [41] is devoted to the study of a traffic flow model on a network composed by an arbitrary number of incoming and outgoing arcs connected together by a node with a buffer. We define the solution to the Riemann problem at the node and we prove existence and well posedness of solutions to the Cauchy problem.

6.1.2. Crowd motion

Participants: Nora Aïssiouene, Christophe Chalons [LJLL, UP7], Régis Duvigneau, Paola Goatin, Matthias Mimault, Massimiliano D. Rosini [ICM, Warsaw University, Poland], Nicolas Seguin [LJLL, UPMC], Monika Twarogowska.

The activity on in pedestrian flow modeling is reinforced by the doctoral thesis of M. Mimault, started in October, and the enrollment of M. Twagorowska on a post-doctoral position.

Concerning crowd motion modeling, we are interested in the optimization of facilities design, in order to maximize pedestrian flow and avoid or limit accidents due to panic situations. To this aim, we are now studying first and second order macroscopic models for crowd movements consisting in one or two scalar conservation law accounting for mass conservation and momentum balance, coupled with an Eikonal equation giving the flux direction depending on the density distribution. From the theoretical point of view, and as a first step, we are studying the problem in one space dimension (for applications, this case corresponds to a crowd moving in a corridor). In collaboration with M. Rosini (supported by the project CROM3, funded by the PHC Polonium 2011), we have established entropy conditions to select physically relevant solutions, and we have constructed explicit solutions for some simple initial data (these results are presented in [40]). We are now studying existence of solutions of the corresponding initial boundary value problem, using the wave-front tracking approach. In this framework, M. Mimault's internship was devoted to develop a MATLAB code based on wave-front tracking to compute the solutions of Hughes' model of pedestrian motion with generalized running cost. This model displays a non-classical dynamic at the splitting point between the two directions of motion. The wave-front tracking scheme provides us with reference solutions to test numerically the convergence of classical finite volume schemes, which do not treat explicitly the dynamics at the turning point (see [66]). The code can be downloaded at the following URL: http://www-sop.inria.fr/members/Paola.Goatin/wft.html

From the numerical point of view, we are implementing some macroscopic models in two space dimensions on triangular meshes on the Num3sis platform. This was partly done by N. El-Khatib (postdoc at Inria from January to August 2011), and is now being completed by M. Twarogowska, with the support of N. Aïssiouene. This will provide a performing numerical tool to solve the related optimization problems arising in the optimization of facilities design, such as the position and size of an obstacle in front of (before) a building exit in order to maximize the outflow through the door and avoid or limit over-compression.

Finally, in collaboration with C. Chalons and N. Seguin, we have generalized the results on conservation laws with local flux constraint obtained in [3], [5] to general flux functions and nonclassical solutions arising for example in pedestrian flow modeling. We first define the constrained Riemann solver and the entropy condition, which singles out the unique admissible solution. We provide a well posedness result based on wave-front tracking approximations and Kruzhkov doubling of variable technique. We then provide the framework to deal with nonclassical solutions and we propose a “front-tracking” finite volume scheme allowing to sharply capture classical and nonclassical discontinuities. Numerical simulations illustrating the Braess paradox are presented as validation of the method. The results are collected in [65].

The above researches were partially funded by the ERC Starting Grant "TRAM3 - Traffic management by macroscopic models".

6.2. Optimum design and control in fluid dynamics and its couplings

In computational sciences for physics and engineering, Computational Fluid Dynamics (CFD) are playing one of the major roles in the scientific community to foster innovative developments of numerical methodologies. Very naturally, our expertise in compressible CFD has led us to give our research on numerical strategies for optimum design a particular, but not exclusive focus on fluids.

The framework of our research aims to contribute to numerical strategies for PDE-constrained multi-objective optimization, with a particular emphasis on CPU-demanding computational applications in which the different criteria to be minimized (or reduced) originate from different physical disciplines that share the same set of design variables. These disciplines are often fluids, as a primary focus, coupled with some other discipline, such as structural mechanics.

Our approach to *competitive optimization* is based on a particular construction of *Nash games*, relying on a *split of territory* in the assignment of individual strategies. A methodology has been proposed for the treatment of two-discipline optimization problems in which one discipline, the primary discipline, is preponderant, or fragile. Then, it is recommended to identify, in a first step, the optimum of this discipline alone using the whole set of design variables. Then, an orthogonal basis is constructed based on the evaluation at convergence of the Hessian matrix of the primary criterion and constraint gradients. This basis is used to split the working design space into two supplementary subspaces to be assigned, in a second step, to two virtual players in competition in an adapted Nash game, devised to reduce a secondary criterion while causing the least degradation to the first. The formulation has been proved to potentially provide a set of Nash equilibrium solutions originating from the original single-discipline optimum point by smooth continuation, thus introducing competition gradually. This approach has been demonstrated over a test-case of aero-structural aircraft wing shape optimization, in which the eigensplit-based optimization reveals clearly superior [38].

While the two-discipline method is currently being applied to various complex physical multi-objective situations (see in particular 6.2.4 , 6.2.5 , 6.2.6), the method has been extended to situations involving more than two objectives when the initial point is Pareto-optimal. Then, a particular convex combination of the criteria is locally stationary, and the two-discipline strategy can be applied using this combination as preponderant criterion, and a particular other criterion as secondary one. Whence, the proposed split of territory produces a continuum of Nash equilibrium points *tangent* to the Pareto set. This theoretical result has been illustrated in the context of a simpler numerical experiment by E. Baratchart during his internship [4], see Fig. 2 .

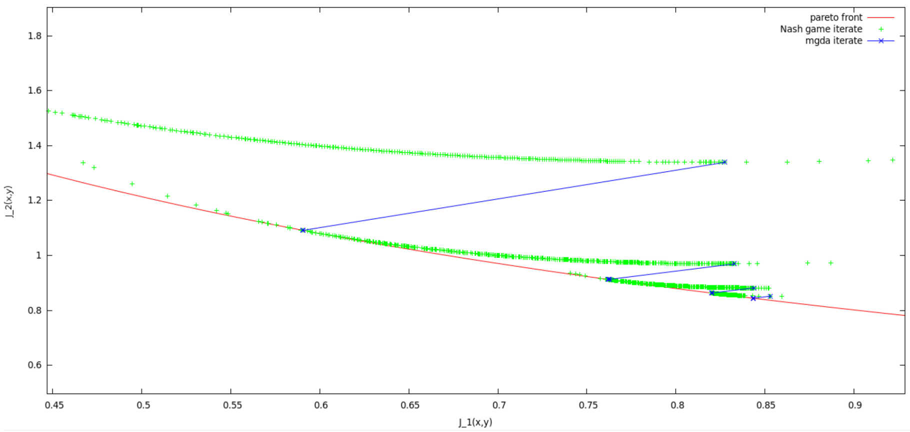


Figure 2. Combination of cooperative and competitive optimization algorithms: in red the Pareto set, in blue MGDA steps directed to the Pareto set, in green steps by Nash games with split of territory tangent to the Pareto set.

Our approach to *cooperative optimization* is based on a result of convex analysis established for a general unconstrained mult-iobjective problem in which all the gradients are assumed to be known. The theorem [39] states that in the convex hull of the gradients, there exists a unique vector of minimal norm, ω; if it is nonzero, the vector ω is a descent direction common to all criteria; otherwise, the current design point is Pareto-optimal. This result led us to generalize the classical steepest-descent algorithm by using the vector ω as search direction. We refer to the new algorithm as the multiple-gradient descent algorithm (MGDA). The MGDA yields to a point on the Pareto set, at which a competitive optimization phase can possibly be launched on the basis of the local eigenstructure of the different Hessian matrices. This general formulation fosters several connected studies detailed in 6.2.1 .

6.2.1. Multiple-Gradient Descent Algorithm (MGDA)

Participants: Jean-Antoine Désidéri, Régis Duvigneau, Matteo Giacomini, Adrien Zerbinati.

6.2.1.1. Theory and numerical experimentation of the MGDA construction

In multi-objective optimization, the knowledge of the Pareto set provides valuable information on the reachable optimal performance. A number of evolutionary strategies (PAES, NSGA-II, etc), have been proposed in the literature and proved to be successful to identify the Pareto set. However, these derivative-free algorithms are very demanding in terms of computational time. Today, in many areas of computational sciences, codes are developed that include the calculation of the gradient, cautiously validated and calibrated.

In the original report [14], and in [39], we have introduced the notion of *Pareto-stationarity*, and given a first proof that it was the natural necessary condition for Pareto-optimality when the objective-functions are locally smooth in some open domain about the design-point. This report has been revised to provide a more rigorous, and extended proof. In particular, in the revised version [14] (version 3, 2012), the number of objective-functions n and the dimension of the design space compare arbitrarily. The objective-functions are assumed to be locally convex.

Additionally, we had established that MGDA converges to Pareto-stationary design-points. This had been confirmed by numerical experiments in which MGDA had been tested over a number of classical multi-objective optimization test-cases, and found successful to converge to Pareto-optimal solutions in situations of either convex or concave Pareto sets. Additionally, MGDA [57] and PAES [69] were found to have complementary merits, making a hybrid method promising.

The method was tested successfully in a domain partition model problem in which the sub-solutions to the Poisson equation are matched at the interfaces by minimization of the integral along the interface of the squared normal-derivative jump. This academic exercise has permitted to illustrate the importance of applying an appropriate scaling to the gradients prior to calculating the descent direction [61] [47]. This has led us to define, a novel form of MGDA, consisting of a direct algorithm [62] based on a Gram-Schmidt orthogonalization conducted with a special normalization. The direct method was found more accurate and more efficient. Subsequently, we proposed two enhancements [63], the first to define the order in which the gradients are introduced in the Gram-Schmidt process uniquely and to interrupt the process as soon as the current estimate of the search direction is proved to satisfy the descent property, and the second to optimally scale the gradients when the Hessians are known, or approximated (e.g. by the BFGS estimate).

6.2.1.2. Meta-model-assisted CFD optimization by MGDA

Using MGDA in a multi objective optimization problem requires the evaluation of a large number of points with regard to criteria, and their gradients. In the particular case of a CFD problems, each point evaluation is very costly since it involves a flow computation, possibly the solution of an adjoint-equation. To alleviate this difficulty, we have proposed to construct meta-models of the functionals of interest (lift, drag, etc) and to calculate approximate gradients by local finite differences. These meta-models are updated throughout the convergence process to the evaluation of the new design points by the high-fidelity model, here the 3D compressible Euler equations.

This variant of MGDA has been tested successfully over a problem of external aerodynamic optimum-shape design of an aircraft wing consisting of reducing wave-drag, and augmenting lift. After only a few cycles of database updates, the Pareto front visibly forms, and this result is achieved at a very moderate computational cost. This variant has been extended successfully to an internal flow optimization problem related to an automobile air-conditioning system and governed by the Navier-Stokes equations [55]. This more difficult problem has been proposed by Renault within the OMD2 ANR project.

6.2.1.3. Exact shape gradients

MGDA has successfully been tested over a two-objective optimization problem governed by two-dimensional elasticity. The deformation of a plate is calculated using an isogeometric approximation (see 6.6) and compliance derived from it. The exact parametric shape gradient is calculated, yielding the gradient of the objective function in two antagonistic situations differing by the loading. Pareto-fronts are thus identified.

6.2.1.4. Perspectives

MGDA offers the possibility to handle in a rational way several objective-functions for which gradients are known or approximated concurrently. This potential opens methodological paths to several themes of interest in high-fidelity simulation-based optimization: optimization of complex systems whose performance is evaluated w.r.t. several criteria originating from different, coupled disciplines; optimization under uncertainties, by introducing sensitivities as additional objectives; optimization of time-dependent systems, such as optimization of flow-control devices that generate a periodic flow (see next subsection), by converting the problem into a multi-point problem by time-discretization of the time and parameter-dependent functional; etc.

6.2.2. Flow control

Participants: Jean-Antoine Désidéri, Régis Duvigneau, Jérémie Labroquère.

Shape optimization methods are not efficient to improve the performance of fluid systems, when the flow is characterized by a strong unsteadiness related to a massive detachment. This is typically the case for the flow around an automotive body or a wing in stall condition. To overcome this difficulty, flow control strategies are developed, that aim at manipulating vortex dynamics by introducing some active actuators, such as periodic blowing/suction jets. In this context, the choice of the control parameters (location, amplitude, frequency) is critical and not straightforward. Therefore, a numerical study is conducted to i) improve the understanding of controlled flows ii) develop a methodology to determine optimal control parameters by coupling the controlled flow simulation with optimization algorithms. Two research axes have been considered :

- the resolution of the unsteady sensitivity equations derived from the state equations, to exhibit the dependency of the flow dynamics with respect to the control ;
- the optimization of control parameters using a statistical metamodel-based strategy[37].

In this perspective, unsteady Reynolds Averaged Navier-Stokes equations are considered, with the Spalart-Allmaras turbulence closure. A numerical model for synthetic jets has been implemented to simulate the actuation[48], based on imposed velocity boundary conditions. Particular developments have then be carried out to include a noise term into Gaussian Process metamodels, which is used to filter errors arising from unsteady simulations/citelabroquere:hal-00742940. First results have demonstrated the feasibility of the proposed method. A systematic assessment of modeling and numerical errors is in progress, for a backward facing step test-case, with the objective of controlling the re-attachment point location.

This activity is conducted in collaboration with the CFD team of Ecole Centrale de Nantes.

6.2.3. Robust design

Participants: Jean-Antoine Désidéri, Régis Duvigneau, Daïgo Maruyama.

This work aims to develop robust design tools for aircraft design w.r.t. aerodynamic performance subject to uncertainties arising from geometrical features and fluctuations of inflow conditions. The robust design process is considered as a multi-objective optimization problem consisting of minimizing statistical quantities such as mean and variance of a cost function, typically the drag coefficient under lift constraint. MGDA is used for this purpose.

At present, analytical test cases have been tested, confirming the validity of our approach to identify the Pareto set.

One aspect of the problem is that the evaluation of these statistics and performing their optimization is very cost demanding. One solution could be, for aerodynamic design, to identify the most important variables to be treated as uncertain, possibly by the ANOVA approach, and construct adequate meta-models.

6.2.4. Aero-structural optimization

Participants: Gérald Carrier [Research Engineer, ONERA/DAAP], Jean-Antoine Désideri, Imane Ghazlane.

In industry, aircraft wings are designed by accounting for several multidisciplinary couplings. Certainly of greatest importance is the coupling, or concurrency, between aerodynamic optimization and structural design. At ONERA, in the former thesis of M. Marcelet, the aerodynamic gradient has been extended to account for (the main terms of) static fluid-structure interaction, commonly referred to as the “aeroelastic gradient”.

In her thesis, I. Ghazlane has extended M. Marcelet’s work to take into account, in the aeroelastic gradient, the terms originating from the differentiation of the wing-structural model. In this development, the wing structure is treated as an equivalent Euler-Bernoulli beam. These formal extensions have been validated by an extensive experimentation. Additionally, special post-processing procedures have been set up to evaluate accurately the various physical contributions to drag. As a result, a realistic aircraft wing optimization has been conducted using a configuration provided by Airbus France as initial design. I. Ghazlane defended successfully her doctoral thesis thesis in December 2012 [34].

Besides, I. Ghazlane has realized a two-objective optimization (drag and mass reduction) via a Nash game using our optimization platform FAMOSA. These results will be included in a common publication on Nash games in preparation.

6.2.5. Sonic boom reduction

Participants: Gérald Carrier [Research Engineer, ONERA/DAAP], Jean-Antoine Désideri, Andrea Minelli, Itham Salah El Din [Research Engineer, ONERA/DAAP].

When an aircraft flies at supersonic speed, it generates at ground level an N-shaped shock structure which can cause serious environmental damage ("sonic boom"). Thus a problem of interest in aerodynamic optimization is to design such an aircraft to reduce the intensity of the sonic boom while maintaining the aerodynamic performance (drag minimization under lift constraint). Andrea Minelli aimed at contributing to this two-discipline optimization problem. In the first part of his work, an inverse problem has been formulated and solved for "shaped sonic boom" and found in excellent agreement with the George-Seebass-Darden theory [68] for the calculation of the Whitham function corresponding to the lowest-boom (axisymmetric) shape. Method and results for more general geometries have been presented internationally in [50].

Besides, aero-acoustic optimizations have been realized successfully by coupling the aerodynamic optimizer (based on Euler calculations by the elsA software) with the sonic-boom computation in a Nash game formulation. These experiments, conducted with our optimization platform FAMOSA, have demonstrated that starting from the shape optimized aerodynamically, one could retrieve smoothly a shape corresponding to nearly-optimal sonic-boom reduction. These results will be included in a common publication on Nash games in preparation.

6.2.6. Helicopter rotor blade optimization in both situations of hovering and forward flight

Participants: Michel Costes [Research Engineer, ONERA/DAAP], Jean-Antoine Désideri, Arnaud Le Pape [Research Engineer, ONERA/DAAP], Enric Roca Leon.

E. Roca Leon is conducting a CIFRE thesis supported by EUROCOPTER (Marignane) at ONERA DAAP. This thesis follows the doctoral thesis of A. Dumont in which the adjoint-equation approach was used to optimize a rotor blade in hovering flight. The goal of this new thesis is to solve a two-objective optimization problem in which the hovering-flight criterion is considered preponderant, but a new criterion that takes into account the forward-flight situation is also introduced, concurrently. The second criterion is the power necessary to maintain the forward motion. The first phase of thesis work has been devoted to the set up of a hierarchy of models from low to high fidelity, in order to calibrate appropriate functional criteria. In the current work, actual two-objective optimizations are conducted via our Nash game approach to competitive optimization with territory splitting based on reduced Hessian diagonalization. A first successful experiment has been realized in which the twist angle along the wing is optimized to reduce the power in forward motion while maintaining sub-optimality of the drag in hover. These results have been accepted for presentation at a forthcoming AIAA Conference, and will also contribute to a common publication on Nash games in preparation.

6.2.7. Optimum design in naval hydrodynamics

Participants: Régis Duvigneau, Louis Blanchard.

Naval hydrodynamics field has recently shown a growing interest for optimum design methods. The computational context is especially complex because it implies unsteady two-phase turbulent flows, with possibly very high Reynolds number (up to 10^9). The use of automated design optimization methods for such problems requires new developments to take into account the large CPU time necessary for each simulation and the specificity of the geometries considered.

In collaboration with GALAAD Project-Team, some developments have been initiated on the geometrical modelling of hull shapes by parametric surfaces. The objective was to be able to modify existing hull shapes by controlling a small number of parameters, that are meaningful for naval architects. We have considered as test-case the bow shape for trawler ships[58]. As a second step, an optimum shape procedure has been set up, based on a metamodel-based optimizer, the developed CAD model and the simulation tool for free-surface flows provided by K-Epsilon company. The objective was to reduce the wave drag of a trawler ship by adding a bow, whose parameters are optimized.

6.3. Optimum design in structural mechanics

6.3.1. Shape Optimization in Multidisciplinary Non-Linear Mechanics

Participants: Aalae Benki, Jean-Antoine Désidéri, Abderrahmane Habbal.

In collaboration with the ArcelorMittal's Center for Research in Automotive and Applications, we study the multidisciplinary shape and parameter design of highly non linear mechanical 2D and 3D structures. We have developed methods adapted to the approximation of Pareto Fronts such as Normal Boundary Intersection NBI and Normalized Normal Constraint Method NNCM. Due to the time consuming cost evaluation, the use of cheap to evaluate surrogate models is mandatory. We have studied the consistency of the approach NBI or NNCM plus surrogates, which turned out to be successful for a broad panel of standard mathematical benchmarks. The coupling is successfully applied to a small scale industrial case, namely the shape optimization of a can bottom vis à vis dome reversal pressure and dome growth criteria. We have then defined a Nash game between criteria where the latter are approximated by the RBF metamodels. First, we validated the computation of a Nash equilibrium for mathematical functions, then we computed Nash equilibria for the small scale industrial case of the shape optimization of the can bottom. In both cases, only arbitrary territory splitting was used. Application to large scale 3D industrial problems, and the study of intelligent territory splitting algorithms is ongoing.

6.3.2. Optimization of Addendum Surfaces in Stamping

Participants: Fatima Zahra Oujebbour, Jean-Antoine Désidéri, Abderrahmane Habbal.

Within the OASIS Consortium (ArcelorMittal, ErDF, Inria, UTC, EURODECISION, ESILV, NECS, Delta-CAD, SCILAB-DIGITEO), Opale Project leads the Optimization task. Our aim is to develop decentralized decision-making algorithms dedicated to find efficient solutions (Pareto optimal) in a complex multidisciplinary framework (forming, stamping, welding non-linear processes, spring-back, vibration, in-function linear processes, crash and fatigue non linear and non differentiable processes) for several (between three and five) criteria. An important difficulty when trying to identify the Pareto Front, even when using adapted methods such the Normal Boundary Intersection, is that the criteria involved (thanks to the high nonlinearity in the mechanical models) exhibit many local optima. So one must use global optimization methods. We have studied the hybrid approach Simulated Annealing with Simultaneous Perturbation SASP for a suite of mathematical test-cases. To envisage the application of our method to the complex CPU time consuming stamping process, we lead an intermediate phase dedicated to the validation of the SASP method for the minimization of the spring-back that follows the stamping of a metal sheet, the design variable being the thickness distribution.

We have successfully applied the NBI approach coupled to the hybrid SA+SPSA minimizer (Simulated Annealing with local search using the Simultaneous Perturbation Stochastic Approximation) to capture the Pareto front of a simple cross stamping of a high performance steel sheet. The use of cubic spline approximation of the costs (spring-back and failure criteria) turned out to be more reliable than e.g. a krigeage method.

6.4. Application of shape and topology design to biology and medicine

6.4.1. Mathematical modeling of dorsal closure DC

Participants: Abderrahmane Habbal, Luis Almeida [University of Nice-Sophia Antipolis], Patrizia Bagnerini [Genova University], Fanny Serman [University of Nice-Sophia Antipolis], Stéphane Noselli [University of Nice-Sophia Antipolis], Glenn Edwards [Duke University].

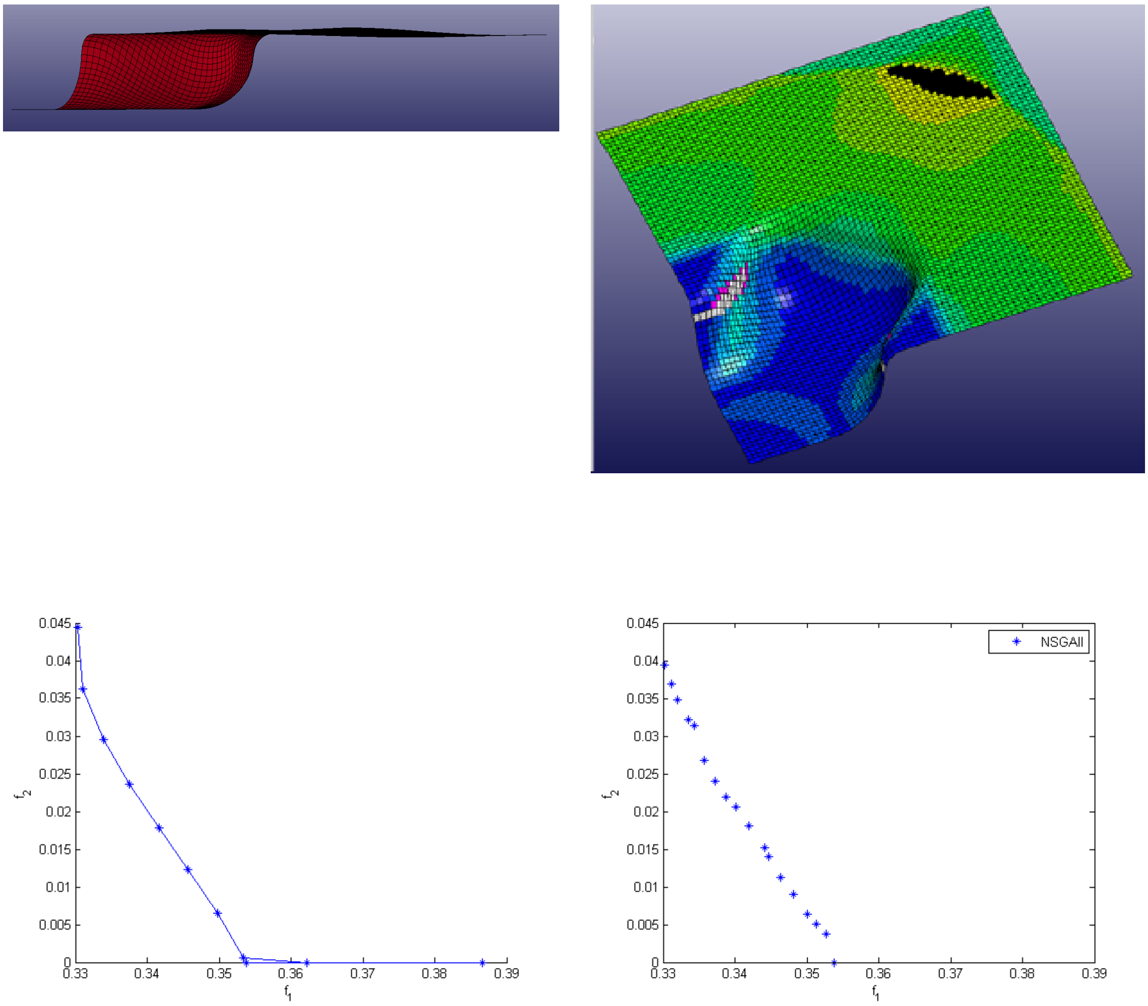


Figure 3. Multiobjective design of the stamping process of a high performance steel sheet. The costs are elastic spring-back (upper-left) and failure (upper-right). The Pareto front obtained by NNCM (lower-left) is compared to a NSGA-II one (lower-right).

A mathematical model for simulation of actin cable contraction, during wound closure for Drosophila embryo, which contains an extra term in addition to the curvature flow is developed. The basic mathematical model introduced and validated in [2] is extended in order to include the non-homogeneous wound healing or non-homogeneous dorsal closure The new model is obtained by adding extra terms that describe the particular process we want to model (lamellipodial crawling, granulation tissue contraction, extension of actin protrusions, epithelial resistance, etc.). We concentrate on the treatment of non-homogeneous forces, i.e. non-constant boundary terms which can be associated with a non-uniform cable, internal pull or zipping force due to the non-uniformity of the biological or physical properties of the boundary cells or of the connective tissue [35].

We also consider a particular yet major aspect of wound healing, namely the one related to the movement of wounded epithelial cell monolayers. The epithelial monolayer cell population, also referred to as cell-sheet, can be seen as a 2 dimensional structure, although it is well known that apical and basal sites play distinctive important roles during the migration, as well as the substrate itself. Immediately after a wound is created, the cells start to move in order to fill in the empty space. This movement, the wound closure, is a highly-coordinated collective behavior yielding a structured cohesive front, the wound leading edge. Even though wound closure involves biochemical and biomechanical processes, still far from being well understood, which are distributed over the whole monolayer, much specific attention was paid to the leading edge evolution, seen as the front of a traveling wave of the cell density function. We show that, for non inhibited wound assays, closure occurs at constant speed of the leading edge, a fact that is commonly shared by biologists and biomathematicians. But we also show that the leading edge may exhibit accelerated profiles, and that when inhibited, then the F-KPP has poor performances in modeling the leading edge dynamics.

6.5. Particular applications of simulation methods

6.5.1. Hermitian interpolation under uncertainties

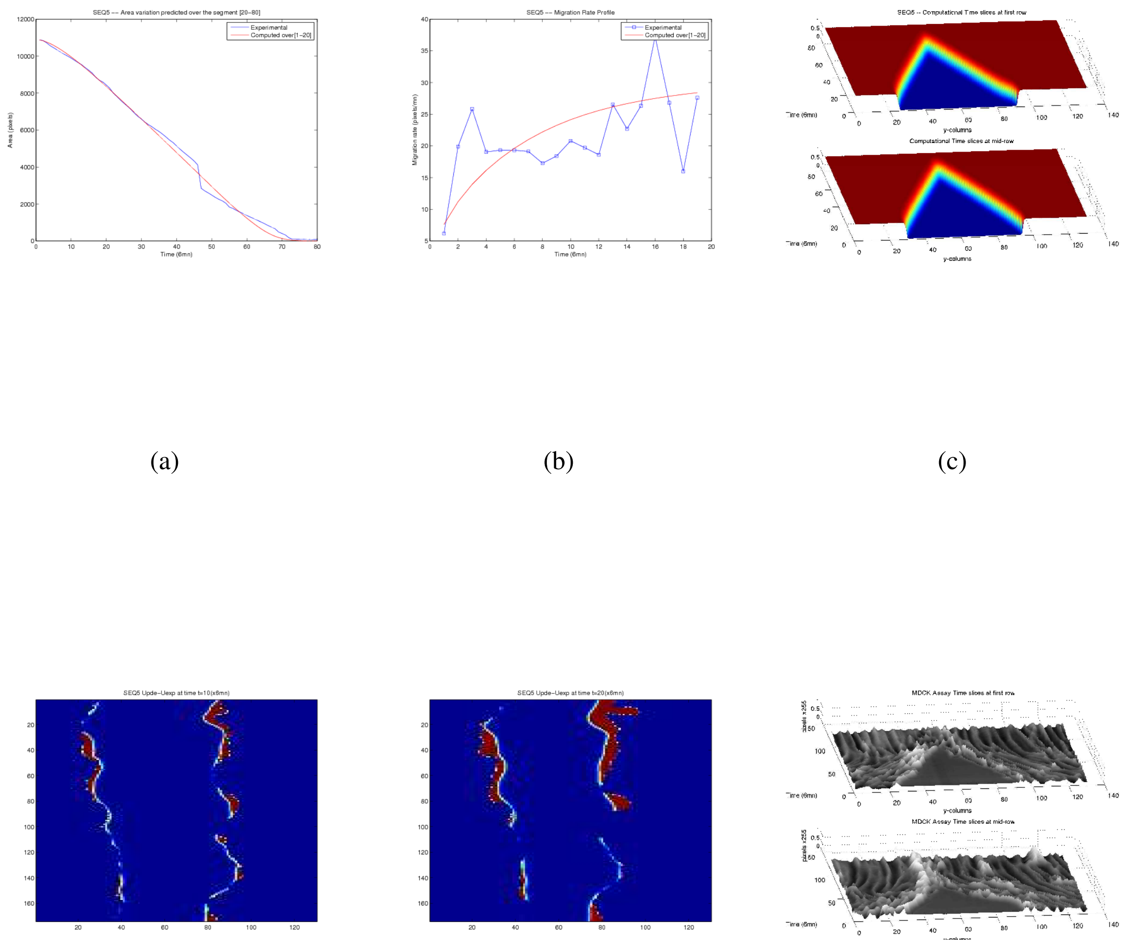
Participants: Jean-Antoine Désideri, Manuel Bompard [Doctoral Student, ONERA/DSNA until December 2011; currently post-doctoral fellow in Toulouse], Jacques Peter [Research Engineer, ONERA/DSNA].

In PDE-constrained global optimization, iterative algorithms are commonly efficiently accelerated by techniques relying on approximate evaluations of the functional to be minimized by an economical, but lower-fidelity model (meta-model), in a so-called Design of Experiment (DoE). Various types of meta-models exist (interpolation polynomials, neural networks, Kriging models, etc). Such meta-models are constructed by pre-calculation of a database of functional values by the costly high-fidelity model. In adjoint-based numerical methods, derivatives of the functional are also available at the same cost, although usually with poorer accuracy. Thus, a question arises : should the derivative information, available but known to be less accurate, be used to construct the meta-model or ignored ? As a first step to investigate this issue, we have considered the case of the Hermitian interpolation of a function of a single variable, when the function values are known exactly, and the derivatives only approximately, assuming a uniform upper bound ε on this approximation is known. The classical notion of best approximation has been revisited in this context, and a criterion introduced to define the best set of interpolation points. This set was identified by either analytical or numerical means. If $n+1$ is the number of interpolation points, it is advantageous to account for the derivative information when $\varepsilon \leq \varepsilon_0$, where ε_0 decreases with n, and this is in favor of piecewise, low-degree Hermitian interpolants. In all our numerical tests, we have found that the distribution of Chebyshev points is always close to optimal, and provides bounded approximants with close-to-least sensitivity to the uncertainties [56].

6.5.2. Mesh qualification

Participants: Jean-Antoine Désideri, Maxime Nguyen, Jacques Peter [Research Engineer, ONERA/DSNA].

M. Nguyen Dinh is conducting a CIFRE thesis at ONERA supported by AIRBUS France. The thesis topic is the qualification of CFD simulations by anisotropic mesh adaption. Methods for refining the 2D or 3D structured mesh by node movement have been examined closely. Secondly, it is investigated how could the local information on the functional gradient $\|dJ/dX\|$ be exploited in a multi-block mesh context. This raises particular questions related to conservation at the interfaces.



(a) (b) (c)

(d) (e) (f)

*Figure 4. **Sequence-5. Computational vs experimental wound evolution.** (a) Time variation of experimental (blue) versus computed (red) wound area (in pixels). (b) Time variation of the experimental (blue-dot) versus computed (red) migration rate (in pixels/mn). (c) Computed 3D XT view at first and mid-rows. (d) (e) (f) Traces of the difference between the experimental segmented and binarized cell-sheet images and the computed ones at different times, respectively 1hour (d), and 2hours (e) after the wounding. (f) Experimental 3D XT view at first and mid-rows.*

Several criteria have been assessed for mesh qualification in the context of inviscid-flow simulation and are currently being extended to the RANS context. These results have been presented internationally in the communication [54] and the publication [44].

6.5.3. Hybrid meshes

Participants: Sébastien Bourasseau, Jean-Antoine Désideri, Jacques Peter [Research Engineer, ONERA/DSNA], Pierre Trontin [Research Engineer, ONERA/DSNA].

S. Bourasseau has started a CIFRE thesis at ONERA supported by SNECMA. The thesis is on mesh adaption in the context of hybrid meshes, that is, made of both structured and unstructured regions. Again, the aim is to exploit at best the function gradient provided by the adjoint-equation approach. Preliminary experiments have been conducted on geometries of stator blade yielding the sensitivities to global shape parameters.

The on-going developments are related to the extension to the hybrid-mesh context of the full shape gradient in a 3D Eulerian flow computation.

6.5.4. Data Completion Problems Solved as Nash Games

Participants: Abderrahmane Habbal, Moez Kallel [University of Tunis].

The Cauchy problem for an elliptic operator is formulated as a two-player Nash game.

- Player (1) is given the known Dirichlet data, and *uses as strategy variable the Neumann condition* prescribed over the inaccessible part of the boundary.
- Player (2) is given the known Neumann data, and *plays with the Dirichlet condition* prescribed over the inaccessible boundary.
- The two players solve in parallel the associated Boundary Value Problems. Their respective objectives involve the *gap between the non used Neumann/Dirichlet known data and the traces of the BVP's solutions* over the accessible boundary, and are *coupled through a difference term*.

We prove the existence of a unique Nash equilibrium, which turns out to be the reconstructed data when the Cauchy problem has a solution. We also prove that the completion algorithm is stable with respect to noise. Many 3D experiments were performed which illustrate the efficiency and stability of our algorithm [42].

6.6. Isogeometric analysis and design

Participants: Louis Blanchard, Régis Duvigneau, Bernard Mourrain [Galaad Project-Team], Gang Xu [Galaad Project-Team].

Design optimization stands at the crossroad of different scientific fields (and related software): Computer-Aided Design (CAD), Computational Fluid Dynamics (CFD) or Computational Structural Dynamics (CSM), parametric optimization. However, these different fields are usually not based on the same geometrical representations. CAD software relies on Splines or NURBS representations, CFD and CSM software uses grid-based geometric descriptions (structured or unstructured), optimization algorithms handle specific shape parameters. Therefore, in conventional approaches, several information transfers occur during the design phase, yielding approximations that can significantly deteriorate the overall efficiency of the design optimization procedure. Moreover, software coupling is often cumbersome in this context.

The isogeometric approach proposes to definitely overcome this difficulty by using CAD standards as a unique representation for all disciplines. The isogeometric analysis consists in developing methods that use NURBS representations for all design tasks:

- the geometry is defined by NURBS surfaces;
- the computation domain is defined by NURBS volumes instead of meshes;
- the solution fields are obtained by using a finite-element approach that uses NURBS basis functions
- the optimizer controls directly NURBS control points.

Using such a unique data structure allows to compute the solution on the exact geometry (not a discretized geometry), obtain a more accurate solution (high-order approximation), reduce spurious numerical sources of noise that deteriorate convergence, avoid data transfers between the software. Moreover, NURBS representations are naturally hierarchical and allows to define multi-level algorithms for solvers as well as optimizers.

In this context, some studies on elliptic problems have been conducted in collaboration with GALAAD Project-Team, such as the development of methods for adaptive parameterization including an a posteriori error estimate[46], [45]. A collaborative work has also been carried out with the Technical University of Kaiserslautern, concerning the computation of shape gradients for linear elasticity problems[59].

POEMS Project-Team

6. New Results

6.1. Numerical methods for time domain wave propagation

6.1.1. Coupling Retarded Potentials and Discontinuous Galerkin Methods for time dependent wave propagation problems

Participant: Patrick Joly.

This topic is developed in collaboration with J. Rodriguez (Santiago de Compostela) in the framework of the contract ADNUMO with AIRBUS. The general objective was to use time-domain integral equations - or retarded potentials - as a tool for contructing transparent boundary conditions for wave problems in unbounded media, by coupling them to an inerior volumic method, namely the Discontinuous Galerkin (DG) method.

Since last year, our new goal is to extend the method proposed in a previous work for DG with central fluxes to the case of upwind fluxes, while preserving most of the good properties of the original method from both theoretical (stability via energy dissipation - instead of energy conservation) and practical points of view. We have designed a method that achieves this goal at the only prize of a small deterioration of the CFL condition. The method has been successfully implemented and the numerical results clearly emphasize the superiority of upwing fluxes for taking into account the convection terms in the linearized Euler equations in aeroacoustics, the privileged application.

At the same time, we have used similar ideas for treating physical boundary conditions involving differential (in time) impedance operators.

6.1.2. Solving the Homogeneous Isotropic Linear Elastodynamics Equations Using Potentials and Finite Elements.

Participants: Aliénor Burel, Marc Duruflé, Patrick Joly.

This topic is the subject of the first part oh th PhD thesis of A. Burel. Its aim is to use the classical theoretical decomposition of the elastodynamic displacement into two potentials referring to the pressure wave and the shear wave, and use it in a numerical context. Last year, a method has been proposed for solving the Dirichlet problem (clamped boundary), successfully analyzed and implemented. For free boundary conditions, we have proposed an original method considereing these boundary conditions as a perturbation of the Dirichlet conditions. The natural adaptation of the variational formulation used in the case of the Dirichlet problems presents nice theoretical properties and leads to satisfactory numerical results for the time harmonic problem. However, the implementation for the time dependent problem reveals severe instability phenomena that seem to be already present in the semi-discrete (in space) problem. In order to understand the cause of these instability (and possibly remedy them) we are currently performing the Kreiss analysis of the half-space problems in the case where Q_1 finite elements are used on the same uniform square grid for both P-waves and S-waves potentials.

6.1.3. Time domain analysis of Maxwell's equations in Lorentz materials

Participants: Maxence Cassier, Lucas Chesnel, Christophe Hazard, Patrick Joly, Valentin Vinoles.

This is the time-domain counterpart of the research done at Poems about frequency domain analysis of metamaterials (see also the section 6.2.7) in the framework of the ANR Project Metamath. One fundamental question is the link between the two problems via the limiting amplitude principle, in particular in the cases where the time harmonic problem fails to be well posed problem in the standard framework. This occurs at certain frequencies (see section) when one considers a transmission problem between a Lorentz material and a standard one.

We are investigating this question from both theoretical and numerical points of view. This is also the object of a collaboration with B. Gralak from the Institut Fresnel in Marseille.

6.1.4. Modeling and numerical simulation of a piano.

Participants: Juliette Chabassier, Marc Duruflé, Sébastien Imperiale, Patrick Joly.

The defense of the PhD thesis of Juliette Chabassier, in March, has marked one of the most spectacular achievements in Poems for the past years, concerning the "complete" physical and mathematical modeling of a grand piano and its computer simulation. This is the result of a quite interdisciplinary work in collaboration with Antoine Chaigne (UME, ENSTA). We refer the reader to the three previous activity reports of Poems for a more detailed description of the scientific developments that have led to the implementation of a parallel code for the simulation of the piano. Using this code, M. Duruflé and J. Chabassier have realized a bank of synthetic sounds that can be used for playing scoreboards (using MIDI files for instance). For more details, and also other additional information about the work, we refer the reader to the Web page : http://modelisation.piano.free.fr.

Although already quite satiafactory, the results obtained by the present version of the code show that there is still room for the improvement of our piano model. One of the ideas consists in improving the quality of the model for the hammers and that is why J. Chabassier and M. Duruflé have proposed an enriched model involving the virations of the hammer's shank. We expect to achieve further progress in this direction through our participation to the ITN (Initial Training Network) European project BATWOMAN (Basic Acoustics Training and Workprogram on Methodologies for Acoustics Network) that has been submitted lst November. This projects regroups 11 partners from 7 different contries and gathers academic people with industrials of the donain, including Steinway.

As a theoretical complement to the numerical developments, we have led a systematic theoretical study of the numerical method used in our code for computing string's vibrations. Our concern was to develop a new implicit time discretization, which is associated with finite element methods in space, in order to reduce numerical dispersion while allowing the use of a large time step. We proposed a new θ-scheme based on different θ-approximations for the flexural and shear terms of the equations, which allows to reduce numerical dispersion while relaxing the stability condition. In particular, we gave some insights of innovative proofs of stability by energy techniques that provide uniform estimates with respect to the CFL number. Theoretical results have been illustrated with numerical experiments corresponding to the simulation of a realistic piano string.

6.1.5. Numerical methods in electromagnetism

Participant: Patrick Ciarlet.

Collaborations with Eric Chung, Tang Fei Yu and Jun Zou (Chinese University of Hong Kong, China), Philippe Ciarlet (City University of Hong Kong, China) Haijun Wu (Nanjing University, China), Stefan Sauter and Corina Simian (Universität Zürich).

The numerical approximation of electromagnetic fields is still a very active branch of research. Below, three lines of work are briefly reported.

Edge finite elements are widely used in 2D/3D electromagnetics, however they approximate very weakly the divergence of the fields. In a recent work with H. Wu & J. Zou, we proposed a method that allows one to approximate the divergence accurately in H^{-s}-norms ($1/2 < s < 1$).

Discontinuous Galerkin finite elements are also very popular, as they allow one to design fast (and accurate) methods to solve PDEs. Jointly with E. Chung and T. F. Yu, we designed a numerical method to solve the 2D/3D time-dependent Maxwell equations, using a high order staggered DG method in the spirit of those introduced by E. Chung and B. Engquist. The method has been analyzed on Cartesian meshes and its generalization to unstructured meshes is under way.

A few years ago, we proposed with Philippe Ciarlet a method to solve some problems in linear elasticity intrinsically. With S. Sauter, C. Simian and Philippe Ciarlet, we studied a similar approach that can be applied to 2D electrostatics. It consists in solving the problem in the electric field directly, using exact or local curl-free approximation of the field. Within this framework, we have been able to derive a general method that allows one to derive intrinsic conforming and non-conforming finite element spaces to compute the electrostatic potential. Generalization to 3D electrostatics and linear elasticity is under way.

6.2. Time-harmonic diffraction problems

6.2.1. Numerical computation of variational integral equation methods

Participants: Marc Lenoir, Nicolas Salles.

The dramatic increase of the efficiency of the variational integral equation methods for the solution of scattering problems must not hide the difficulties remaining for an accurate numerical computation of some influence coefficients, especially when the panels are close and almost parallel.

The formulas have been extended to double layer potentials and, for self influence coefficients, to affine basis functions. Their efficiency for the solution of Maxwell equations has been proved in the framework of a collaboration with CERFACS.

6.2.2. Formulation and Fast Evaluation of the Multipole Expansions of the Elastic Half-Space Fundamental Solutions

Participants: Marc Bonnet, Stéphanie Chaillat.

The use of the elastodynamic half-space Green's tensor in the FM-BEM is a very promising avenue for enhancing the computational performances of 3D BEM applied to analyses arising from e.g. soil-structure interaction or seismology. This ongoing work is concerned with a formulation and computation algorithm for the elastodynamic Green's tensor for the traction-free half-space allowing its use within a Fast Multipole Boundary Element Method (FM-BEM). Due to the implicit satisfaction of the traction-free boundary condition achieved by the Green's tensor, discretization of (parts of) the free surface is no longer required. Unlike the full-space fundamental solution, the elastodynamic half-space Green's tensor cannot be expressed in terms of usual kernels such as e^{ikr}/r or $1/r$. Its multipole expansion thus cannot be deduced from known expansions, and is formulated in this work using a spatial two-dimensional Fourier transform approach. The latter achieves the separation of variables which is required by the FMM. To address the critical need of an efficient quadrature for the 2D Fourier integral, whose singular and oscillatory character precludes using usual (e.g. Gaussian) rules, generalized Gaussian quadrature rules have been used instead. The latter were generated by tailoring for the present needs the methodology of Rokhlin's group. Numerical tests have been conducted to demonstrate the accuracy and numerical efficiency of the proposed FMM. In particular, a complexity significantly lower than that of the non-multipole version was shown to be achieved. A full FM-BEM based on the proposed acceleration method for the half-space Green's tensor is currently under way.

6.2.3. Domain decomposition methods for time harmonic wave propagation

Participants: Francis Collino, Patrick Joly, Mathieu Lecouvez.

This work is motivated by a collaboration with the CEA-CESTA (B. Stupfel) through the PhD thesis of M. Lecouvez that has started at the beginning of the year.

We are interested in the diffraction of time harmonic electromagnetic waves by perfectly conducting objects covered by multi-layered (possibly thin) dielectric coatings. This problem is computationally hard when the size of the object is large (typically 100 times larger) with respect to the incident wavelength. In such a situation is to use a domain decomposition method in which each layer would contitute a subdomain. More precisely, we want to use a non overlaping iterative domain decomposition method based on the use of Robin type transmission conditions, a subject to which people at Poems gave substantial contributions in the 90's through the works of Collino, Desprès, and Joly.

The novelty of our approach consists in using new transmission conditions using some specific impedance operators in order to improve the convergence properties of the method (with respect to more standard Robin conditions). Provided that such operators have appropriate functional analytic properties, the theory shows that one achieves geometric convergence (in opposition the the slow algebraic convergence obtained with standard methods). These properties prevent the use of local impedance operator, a choice that was commonly done for the quest of optimized transmission conditions (following for instance the works of Gander, Japhet, Nataf). We propose a solution that uses nonlocal integral operators using appropriate Riesz potentials. To overcome the disadvantage of dealing with completely nonlocal operators, we suggest to work with truncated kernels, i.e. with operators of the form (Γ represents one interface)

$$u(x) \quad \longrightarrow \quad \int_\Gamma K\,(\,|x-y|\,)\ \chi\left(\frac{|x-y|}{\lambda}\right)\ u(y)\ d\sigma(y)$$

where $K(|x|)$ is an appropriate singlar kernel (typically $K(|x|) = |x|^{-\gamma}$) and $\chi(\rho)$ an adequate smooth cut-off function. Playing with a few parameters such as the size of the support of χ, we expect to achieve an optimal compromise between the reduction of the number of iterations of the method and the cost of each iteration.

6.2.4. Time harmonic aeroacoustics

Participants: Anne-Sophie Bonnet-Ben Dhia, Jean-François Mercier.

We are still working on the numerical simulation of the acoustic radiation and scattering in presence of a mean flow. This is the object of the ANR project AEROSON, in collaboration with Florence Millot and Sébastien Pernet at CERFACS, Nolwenn Balin at EADS and Vincent Pagneux at the Laboratoire d'Acoustique de l'Université du Maine. Let us recall that our method combines, a Finite Element resolution of the augmented Galbrun equation and of the coupled vorticity transport equation, and the use of Perfectly Matched Layers (PML) to bound the computational domain. The main recent improvements concern the test of the method in presence of unstable modes.

When determining the aeroacoustics modes propagating in a flow, unstable modes exist for certain types of flows: when an inflection point exists in the velocity profile and when the shear in this point is strong enough. Such modes grow exponentially in space. Up to recently, our numerical simulations have been performed for stable flows. We have tested the behavior of PML in the presence of unstable modes, which usually convert a propagating field in a decaying field. Therefore we do not have a theoretical framework to characterize the behavior of PML in the presence of spatially growing modes but the various conducted numerical tests have shown that our numerical method is still able to select the outgoing solution, even in the presence of instabilities, if the attenuation in the PML is strong enough.

6.2.5. Multiple scattering in a duct

Participant: Jean-François Mercier.

This topis is developed in collaboration with Agnès Maurel (Langevin Institute ESPCI).
The objective of this work, part of the ANR Procomedia, is to develop analytical methods to describe the propagation of acoustic waves in 2D waveguides containing penetrable inclusions. Scatterers of arbitrary shape with a contrast in both density and sound speed are considered. A modal approach is adopted, in which the wave equation is projected onto the transverse modes of the homogeneous guide. For each mode a 1D wave equation is obtained with a source term which characterizes the scatterers and couples modes together. In weak scattering regime (small scatterers or low contrasts or low frequency), the Born approximation is used to solve analytically this family of coupled ODE. This gives an explicit prediction for the scatterered field, in particular the reflection and transmission coefficients are obtained in two cases of interest: periodically or randomly distributed scatterers. In both cases, expressions similar to those in free space (available only for low frequencies) are obtained without frequency limit, thanks to the presence of a shape factor sensitive to the geometry of the scatterers at high frequencies.

Recently the obtained analytical expressions have been exploited to develop a very simple imaging method in a heterogeneous waveguide. Measurements of low-frequency reflection and transmission allow to find the position of the object while the higher frequency measurements give access to the shape and to the physical characteristics of the scatterers. The results are good in the case of low contrast and small scatterers, for which the Born approximation is perfectly valid.

6.2.6. Localization in perturbed periodic metamaterials

Participant: Jean-François Mercier.

This topis is developed in collaboration with Agnès Maurel, Abdelwaheb Ourir (Langevin Institute ESPCI) and Vincent Pagneux (LAUM).

The aim of this work, part of the ANR Procomedia, is to study the propagation of electromagnetic waves through 1D perturbed periodic media. The attenuation length in a medium consisting of alternating materials of optical indices $n_1 > 0$ and $n_2 < 0$ (metamaterials) is determined. When such medium is randomly disturbed, the localization properties differ significantly from those obtained in a classical disturbed medium: in the homogeneous case $n_1 = n_2$, a random perturbation of the indices induces the Anderson localization with a strong field attenuation. In contrast, in the case $n_1 = -n_2$, it was recently shown that the introduction of disorder on the permittivities ϵ_1 and ϵ_2 gave rise to an "anomaly", the suppression of the Anderson localization. This anomaly results in a significant increase of the attenuation length l_N for large sample sizes N.

We have made two improvements to existing works: simple analytical expressions of the attenuation length have been determined, valid over a wide range of frequencies and of number of layers. In addition we considered realistic metamaterials by taking into account disorder in both the permittivity and the permeability μ. When only the permeability is disturbed (or only the permittivity), our analytical expression can explain the transition to the abnormal behavior when the number of layers increases. Furthermore we show that the anomaly is strongly affected when disturbances in permeability and permittivity are jointly considered: the coupling of the two effects is capable of reseting the usual localization.

6.2.7. Modeling of meta-materials in electromagnetism

Participants: Anne-Sophie Bonnet-Ben Dhia, Camille Carvalho, Patrick Ciarlet, Lucas Chesnel.

This topis is developed in collaboration with Eric Chung (Chinese Univ. of Hong Kong) and Xavier Claeys (Paris VI).

Meta-materials can be seen as particular media whose dielectric and/or magnetic constant are negative, at least for a certain range of frequencies. This type of behavior can be obtained, for instance, with particular periodic structures. Of special interest is the transmission of an electromagnetic wave between two media with opposite sign dielectric and/or magnetic constants. As a matter of fact, applied mathematicians have to address challenging issues, both from the theoretical and the discretization points of view. The year 2012 saw the completion of Lucas Chesnel PhD thesis. We present below the main results obtained these last three years. The first topic we considered a few years ago was: when is the (simplified) scalar model well-posed in the classical H^1 framework? It turned out this issue could be solved with the help of the so-called T-coercivity framework. While numerically, we proved that the (simplified) scalar model could be solved efficiently by the most "naive" discretization, still using T-coercivity. Recently, we have been able to provide sharp conditions for the T-coercivity to hold in general 2D and 3D geometries, which involve explicit estimates in simplified geometries together with localization arguments. We then analyzed the discretization of the scalar problem with a classical, H^1 conforming, finite element method, and proved the convergence under the same sharp conditions. We also showed that the problem can be solved with the help of a Discontinuous Galerkin discretization, which allows one to approximate both the field and its gradient (with E. Chung).

As a second topic, we investigated the case of a 2D corner which can be ill-posed (in the classical H^1 framework). Using the Mellin transform, we showed that a radiation condition at the corner has to be imposed to restore well-posedness (with X. Claeys). Indeed there exists a wave which takes an infinite time to reach the corner: this "black hole" phenomenon is observed in other situations (elastic wedges for example). We proposed a numerical approach to approximate the solution which consists in adding some PMLs in the neighbourhood of the corner.

Last, we studied the transmission problem in a purely 3D electromagnetic setting from a theoretical point of view. We proved that the Maxwell problem is well-posed if and only if the two associated scalar problems (with Dirichlet and Neumann boundary conditions) are well-posed. Of course, these scalar problems involves sign-changing coefficients but they can be studied using simple scalar T-coercivity approach.
C. Carvalho started her PhD thesis this fall in the continuation of these works.

6.2.8. Numerical MicroLocal Analysis

Participants: Jean-David Benamou, Francis Collino, Simon Marmorat.

Numerical microlocal analysis of harmonic wavefields is based on a family of linear filters using Bessel functions and applied to wave data collected on a circle of fixed radius r_0 around the observation point x_0 where we want to estimate the Geometric Optics/ High Frequency components. The data can easily be reconstructed from more conventional line array or grid geometry. The output is an angular function presenting picks of amplitudes in the direction angles of rays.

The original NMLA algorithm relied on a local plane wave assumption for the data. For arbitrary waves, it meant linearization errors and accuracy limitations. Also, only the directions of the (multiple) rays are recovered but the traveltime and amplitudes are not reliably computed. We recently introduced a new "impedant" observable which allows to prove a stability theorem. Numerical results confirm that the new NMLA filter is robust to random and correlated noise.

Using asymptotic expansion on NMLA filtered point sources data, we designed a correction method for the angle which also estimates the wavefront curvature. It can be used to correct the linearization errors mentioned above and provides a second order correction in the Taylor approximation of the traveltime.

The parameters of the method (size of observation circle, discretization) are automatically optimized and a posteriori quantitative error on angles and curvature are available. Numerical studies validate the stability result and confirm the superior accuracy of the curvature corrected NMLA version over image processing methods.

When some bandwith is available we can also compute the traveltime. The amplitude remains polluted by phase errors. Its determination is still open.

6.3. Absorbing boundary conditions and absorbing layers

6.3.1. Evolution problems in perturbed infinite periodic media

Participant: Sonia Fliss.

For parabolic problems set in locally perturbed periodic media, we have developed an approach to determine the time-domain DtN operator. The principle is to apply the Laplace Transform in time to the equation and use the construction of the DtN operator for stationary equations. The main difficulty is the computation of the inverse of the Laplace Transform, more precisely to understand how to deal with the unbounded interval of integration and the choice of the discretization of the laplace variable. To deal with the first difficulty for waveguide problem, we have studied the asymptotic behavior of the DtN operator in the laplace domain when the laplace variable tends to $p_0 \pm \infty$. To deal with the second difficulty, we have used the Z-Transformation and its properties. The numerical study is still in progress. This work enters in the framework of the ANR PRoject MicroWave (Sonia Fliss is an external collaborator), in collaboration with Karim Ramdani (Institut Elie Cartan de Nancy, UMR CNRS 7502), Christophe Besse and Ingrid Violet (Laboratoire Paul Painlevé, UMR CNRS 8524).

6.3.2. New transparent boundary conditions for time harmonic acoustic problem in anisotropic media

Participants: Anne-Sophie Bonnet-Ben Dhia, Sonia Fliss, Antoine Tonnoir.

This topis is developed in collaboration with Vahan Baronian (CEA). Many industrial applications require to check the quality of structures such as plates, for instance in aircraft design. A common way to inspect structures is to propagate ultrasonic waves and detect from the experimental results the presence or not of a defect or a crack. However, in aeronautics, structures are often complex media like anisotropic elastic plates for which the interpretation of this results is complicated. Therefore, efficient and accurate numerical methods of simulation are required. In our work, we want to study the diffraction of a time harmonic wave by a bounded defect in an anisotropic elastic media. In order to study the diffraction properties of the defect, we consider it in as infinite. Since the defect has an arbitrary geometry, we want to use a finite element method in a box that surround the defect. On the boundary of this artificial box, we need to find transparent conditions to simulate an infinite domain.

- We first have considered waveguides. The transparent boundary conditions are often written by using the so-called Dirichlet-to-Neumann maps which can be expressed thanks to a modal decomposition. However, classical iterative method does not converge necessarily. In this work, we introduce a new Dirichlet-to-Neumann operator which links the trace of the solution on a section of the waveguide to the normal trace on a different one. This operator can also be expressed analytically via a modal decomposition. Its main advantage is that, because of the overlapping, it becomes compact and this is exactly why we think an iterative resolution has more chance to converge. Other advantages will appear with the elasticity application. Indeed, in the formulation of the transparent boundary condition without overlapping, appears a lagrange multiplier which makes the resolution more costly. This additional unknown will be avoided with an overlap. For now, the theory is done for the scalar acoustic waveguide and the method has been implemented in the Melina code for the acoustic and the elastic case. The redaction of an article is in progress.

  item We then have studied scattering problem in locally perturbed anisotropic plate. The classical methods to derive transparent boundary conditions for acoustic isotropic media are based on the Green function (boundary integral formulation) or Fourier series (to determine DtN operator set on an artificial circle boundary). However, they cannot be extended for anisotropic elastic problems. Using a constructive method to determine transparent boundary conditions for periodic media developped in the laboratory, we were able to propose new exact boundary conditions which are adapted to anisotropic media and for which iterative method could converge rapidly. The numerical study is in progress for acoustic isotropic problem.

6.4. Waveguides, resonances, and scattering theory

6.4.1. Localized modes in periodic waveguides

Participants: Anne-Sophie Bonnet-Ben Dhia, Bérangère Delourme, Sonia Fliss, Sergei Nazarov, Elizaveta Vasilevskaia.

The general objective is the study of localized modes in locally perturbed periodic media. We investigate the existence theory of such modes as well as their numerical computations. We can distinguish two types of problems.

Numerical computation of guided modes in periodic media with line defects. We are interested in the propagation of guided modes that propagate in the direction of the line defect (which is parallel to one of the periodicity directions of the unperturbed medium) and decrease exponentially in the transverse directions. We aim at computing these modes and their dispersion relation. Last year, we developed a method based on the use of the DtN approach introduced in the PhD thesis of S. Fliss and the resolution of "operator pencil" eigenvalue problems. This year, in collaboration with Kersten Schmidt, we have made a numerical comparison of this new method with the more standard supercell method.

Existence of localized modes in closed periodic waveguides. We consider a propagation medium which is infinite and periodic in one space dimension and bounded in the transverse ones. We investigate the question of the influence of a local defect on the existence of localized modes. Once again this reduces to a selfadjoint eigenvalue problem in an unbounded domain.

The first problem that we studied is in the framework of the PostDoc of Bérangère Delourme. We have considered general locally perturbed periodic media for which we focus on determining sufficient conditions on the periodic media or the local defect so that it exists at least one eigenvalue below the essential spectrum of the underlying perfectly periodic operator. These sufficient conditions are based on Min-Max theory and an appropriate choice of test functions. We were able to validate these existence conditions thanks to the numerical method based on the use of DtN operators. For situations where the periodic "reference medium" is closed to a simple "limit medium" fo which all calculations can be made by hand, we show that these conditions could be really simple and explicit using perturbation theory and asymptotic expansions of the eigenvalues. We are investigating now the extension of this approach to sufficient conditions for existence of guided modes inside the essential spectrum.
The second case, that is investigated in the framework of the PhD thesis of E. Valisevskaia, is the case where the propagation medium is a thin structure (the thinness being characterized by the parameter ε) whose limit is a periodic graph. This is for instance the case of a symmetric ladder as illustrated by figure . If Neumann boundary conditions are considered, it is well known (see in particular the works by Exner, Kuchment) the the limit model when εtends to is the Helmholtz equation on the graph (1D Helmholtz equations on each branch competed by continuity and Kirchoff transimission conditions at each node) . For this limit problem, the underlying operator does not present any spectral gaps but can be written, due to the symmetry of the problem, as the sum of two operators, each of which having an infinity of spectral gaps. This allows us to look for eigenvalues in these spectral gaps, induced by symmetric and localized perturbations of the limit graph model. This can de done for instance by modifying (symetrically) the Kirchoff conditions on two symmetric nodes of the graph. In the limit process mentionned above, this would correspond to modifying the width of the rung that joins these two points in the original problem. First existence results have been obtained in this direction. In a further step, one can expect, by asymptotic analysis, to get corresponding existence results for the original problem, at least for ε small enough.

6.4.2. A new approach for the numerical computation of non linear modes of vibrating systems

Participants: Anne-Sophie Bonnet-Ben Dhia, Jean-François Mercier.

A collaboration with Cyril Touzé and François Blanc (Unité de Mécanique, ENSTA). The simulation of vibrations of large amplitude of thin plates or shells requires the expensive solution of a non-linear finite element model. The main objective of the proposed study is to develop a reliable numerical method which reduces drastically the number of degrees of freedom. The main idea is the use of the so-called non-linear modes to project the dynamics on invariant subspaces, in order to generate accurate reduced-order models. Cyril Touzé from the Unité de Mécanique of ENSTA has derived an asymptotic method of calculation of the non-linear modes for both conservative and damped systems. But the asymptotically computed solution remains accurate only for moderate amplitudes. This motivates the present study which consists in developing a numerical method for the computation of the non-linear modes, without any asymptotic assumption. This is the object of a collaboration with Cyril Touzé, and new results have been obtained during the post-doc of François Blanc in the Unité de Mécanique of ENSTA. The partial differential equations defining the invariant manifold of the non-linear mode are seen as a vectorial transport problem : the variables are the amplitude and the phase (a, φ) where the phase φ plays the role of the time. In the case of conservative systems, a finite difference scheme is used and an iterative algorithm is written, to take into account the 2π-periodicity in φ which is seen as a constraint. An adjoint state approach has been introduced to evaluate the gradient of the coast function. The method has been validated in a simple example with two degrees of freedom. Good agreement with an alternative method, the continuation of periodic solutions method, has been found. Currently the method is extended to the case of damped systems. The main difficulty is that, due to a change of variables, the 2π-periodicity does not hold anymore and new constraints more complicated to implement must be considered.

6.4.3. Harmonic wave propagation in locally perturbed periodic waveguide

Participants: Sonia Fliss, Patrick Joly.

We work on the expression and the asymptotic behaviour of the Green function for time harmonic wave equation in two-dimensional periodic waveguide. This enables us to define a radiation condition and show well-posedness of the Helmholtz equation set in a periodic waveguide. The redaction of an article is ongoing. This analysis is one of the main tool to solve inverse problems in locally perturbed periodic waveguide (see section 6.6.1) when the data are far field measurements of scattering problems.
One challenging perspective of this work is to extend these results to periodic problems in free space.

6.4.4. Finite element approximation of modes of elastic waveguides immersed in an infinite fluid

Participants: Anne-Sophie Bonnet-Ben Dhia, Cédric Doucet, Christophe hazard.

This work is developped in collaboration with Vahan Baronian (CEA). We are developping numerical tools to simulate ultrasonic non-destructive testing in elastic waveguides. This particular topic aims at finding an efficient way of coupling semi-analytical finite element methods and perfectly matched layers (PMLs) to compute modes of elastic waveguides embedded in an infinite fluid.
During our numerical investigations, we noticed that the semi-analytical mixed finite element formulation proposed in the PhD thesis of V. Baronian may lead to the computation of spurious modes. We overcame this problem in the following way: instead of approximating components of stress tensors by means of first-order finite elements of class $\mathcal{C}^0$, we decided to use zeroth-order discontinuous ones. This simple modification seems not only to stabilize the discretization step, but also to approximate modes more accurately in comparison with the classical semi-analytical finite element formulation. Last but not least, we observed a meaningful improvement of the approximation of the continuous spectrum of stretched operators related to PMLs. Besides, previous results (in the PhD thesis of B. Goursaud) about the best way of designing PMLs to simulate wave propagation in open acoustic waveguides have been confirmed by our numerical experiments on immersed elastic structures.
Further investigations need to be carried out to explain these phenomena. Especially, a theoretical analysis still remains to be done.

6.5. Asymptotic methods and approximate models

6.5.1. Effective boundary conditions for thin periodic coatings

Participants: Mathieu Chamaillard, Patrick Joly.

This topic is the object of a collaboration with Houssem Haddar. We are interested in the construction of "equivalent" boundary condition for the diffraction of waves by an obstacle with smooth boundary Γ covered with a thin coating of width δ whose physical characteristics vary "periodically along Γ with a period proportional to the small parameter δ. For a general boundary Γ, the notion of periodicity is ambiguous: we have chosen to define the coating as the image, or the deformation, by a smooth mapping of a flat layer of width delta (the reference configuration) that preserves the normals, which appears consistent with a manufactoring process. The electromagnetic parameters in the coating are then defined as the images through Φ_Γ of periodic functions in the reference configuration.

We have first considered the case of the scalar wave equation. Using an asymptotic analysis in δ, which combines homogenization and matched asymptotic expansions, we have been able to establish a second order boundary condition of the form

$$\partial_\nu u + \left(\delta\, B^1_\Gamma + \delta^2\, B^2_\Gamma\right) u = 0$$

where B_Γ^1 and B_Γ^2 are second order tangential differential operators along Γ whose coefficients depend on both the geometrical characteristics of Γ (through the curvature tensor) and the material properties of the coating, through the resolution of particular cell problems in the flat reference configuration. When the coating is homogeneous, we have checked that one recovers the well known second order thin layer condition. This new condition is expected to provide $O(\delta^3)$ accuracy. Its implementation and its rigorous analysis (error estimates) are ongoing.

6.5.2. Thin Layers in Isotropic Elastodynamics

Participants: Marc Bonnet, Aliénor Burel, Patrick Joly.

This research is developed in the framework the numerical modeling of non-destructive testing experiments using ultrasonic waves. Most realistis propagation media involves thin layers of resin (typically for gluing together different homogeneous media), which are, until now, difficult to take into account numerically, the principal issue being the very small space step needed for meshing such a thin layer. An idea to get rid of this complication is to use asymptotic analysis in order to establish effective transmission conditions. We have studied the simple model problem in two dimensions, with an infinite flat layer of thickness ε. Using a formal approach based on a scaling inside the layer and an power series expansion in ε solution as a polynomial in ε, we have established first and second order conditions. Energy techniques parmit to guaranty the stability of our approximation.

6.5.3. Homogenization and metamaterials

Participants: Sonia Fliss, Patrick Joly, Valentin Vinoles.

This topic is developped in collaboration with Xavier Claeys (LJLL, Paris VI).
The mathematical modeling of electromagnetic metamaterials and the homogenization theory are intimately related because metamaterials are precisely constructed by a periodic assembly of small microstructures involving dielectric materials presenting a high contrast with respect to a reference medium. As a consequence, each microstructure behaves as a resonator which induces surprising properties to the effective or homogenized material such as negative permittivity and / or permeability at certain frequencies. The relevant theoretical approach to this question is the non standard (or high contrast) homogenization theory developed in particular in France by G. Bouchitté.

In the framework of the ANR Metamath, we wish to deepen this question by looking carefully at the treatment of boundaries and interfaces that are generally poorly taken into account by the first order homogenization. This is developed in collaboration with X. Claeys (Paris VI).

This question is already relevant for standard homogenization for which taking into account the presence of a boundary induces a loss of accuracy due to the inadequation of the standard homogenization approach to take into account the boundary layers induced by the boundary. Our objective is to construct approximate effective boundary conditions that would restore the desired accuracy.

With the PhD thesis of V. Vinoles, we aim at extending the previous approach to the treatment of metamaterials via high contrast homogenization. In particular, we intend to treat the challenging question of interfaces between metamaterials and standard materials (see also sections).

6.5.4. Asymptotic analysis and negative materials

Participants: Lucas Chesnel, Sergei Nazarov.

This topic is developped in collaboration with Xavier Claeys (LJLL, Paris VI) and S.A. Nazarov (IPME RAS, St Petersburg, Russia).

One of the applications of negative materials (metals at optical frequencies or negative metamaterials) is the construction of subwavelength cavities. In this kind of application, the idea is to use the following result: an inclusion of a negative material in a positive material changes radically the spectrum of the Maxwell's operators. We demonstrated this result for the scalar operator in a configuration where a positive material contains a small negative inclusion whose size tends to zero. As a second topic, we proved an instability result for a configuration where the interface between the positive and the negative material has a rounded corner. It appears that the solution depends critically on the value of the rounding parameter and does not converge when the rounded corner tends to the actual corner. We also studied the spectrum of the scalar operator in this configuration. This spectrum does not converge but seems (for the moment, the proof is not complete) to oscillate like $\ln \delta$ where $\delta \to 0$ is the rounding parameter.

6.5.5. Modelling of non-homogeneous lossy coaxial cable for time domain simulation.

Participants: Geoffrey Beck, Sébastien Imperiale, Patrick Joly, Martina Novelinkova.

This topic, initiated at the end of the PhD thesis of S. Imperiale, has been the subject of the internship of M. Novelinkova and is the subject of the PhD thesis of G. Bech which started in October.

We investigate the question of the electromagnetic propagation in thin electric cables from a mathematical point of view via an asymptotic analysis with respect to the (small) transverse dimension of the cable: as it has been done in the past in mechanics for the beam theory from 3D elasticity, we use such an approach for deriving simplified effective 1D models from 3D Maxwell's equations. Doing so, we have been able to derive a generalized telegraphist's equation, a 1D wave equation with additional time convolution terms that results from the conjugated effect of electromagnetic losses and heterogeneity of the cross section. This new model has been fully justified through error estimates. We are currently working on a higher order generalized telegraphist's equation that would include dispersive effects through nonlocal capacity and inductance operators.

From the pratical point of view, a code that computes the coefficients (including the convolution kernel) of the effective model and solves the generalized telegraphist's equation has been implemented. It has been exploited to measure the presence of localized defects on the propagation of electromagnetic waves. This application has been motivated by the ANR project SODDA, in collaboration with CEA-LETI, about the non destructive trsting of networks of electric cables (a subject that we are investigating in collaboration with M. Sorine from Inria Rocquencourt).

6.5.6. Elastic wave propagation in strongly heterogeneous media

Participants: Patrick Joly, Simon Marmorat.

This subject enters our long term collaboration with CEA-LIST on the development on numerical methods for time-domain non destructive testing experiments using ultra-sounds. This is also the subject of the PhD thesis of Simon Marmorat. Our objective is to develop an efficient numerical approach for the propagation of elastic waves in a medium which is made of many small inclusions / heterogeneities embedded in a smooth (or piecewise smooth) background medium, without any particular assumption (such as periodicity) on the spatial distribution of these heterogeneities. Our idea is to exploit the smallness of the inclusions (with respect to the wavelength in the background medium) to derive a simplified approximate model in which each inclusion would be described by very few parameters (functions of time) coupled to the displacement field in background medium for which we could use a computational mesh that ignores the presence of the heterogeneities. For deriving such a model. we intend to use and adapt the asymptotic methods previously developed at Poems (such as matched asymptotic expansions).

6.5.7. Multiple scattering by small scatterers

Participants: Maxence Cassier, Christophe Hazard.

We consider the scattering of an acoustic time-harmonic wave by an arbitrary number of sound-soft obstacles located in a homogeneous medium. When the size of the obstacles is small compared with the wavelength, the numerical simulation of such a problem by classical methods (e.g., integral equation techniques or methods based on a Dirichlet-to Neumann map) can become highly time-consuming, particularly when the number of scatterers is large. In this case, the use of an asymptotic model may reduce considerably the numerical cost. Such a model was introduced by Foldy and Lax in the middle of the last century to study multiple isotropic scattering in a medium which contains randomly distributed small scatterers. Their asymptotic model is based on the fact that the scattered wave can be approximated by a wave emitted by point sources placed at the centers of the scatterers; the amplitudes of the sources are calculated by solving a linear system which represents the interactions between the scatterers. Nowadays, the FoldyñLax model is still used in numerous physical and numerical applications to approximate the scattered wave in a deterministic media. But to the best of our knowledge, there was no mathematical justification of this asymptotic model. We have proposed such a justification which provides local error estimates for the two-dimensional problem in the case of circular obstacles. An article on this subject has been accepted and will be published in Wave Motion in January 2013.

6.6. Imaging and inverse problems

6.6.1. Sampling methods in waveguides

Participants: Laurent Bourgeois, Anne-Claire Egloffe, Sonia Fliss, Mathieu Guenel, Eric Lunéville.

First, we have adapted the modal formulation of sampling methods (Linear Sampling Method and Factorization Method) to the case of a periodic waveguide in the acoustic case. This study is based on the analysis of the far field of scattering solutions in cylindrical waveguides, in particular for the fundamental solution, which enables us to obtain a far field formulation of sampling methods, and then a modal formulation of such methods. The aim of the inverse problem is to retrieve a defect from the scattered fields which correspond to the incident fields formed by the Floquet modes. The corresponding numerical implementation was the subject of the Master internship of Mathieu Guenel who obtained some first promising results.
Secondly, going back to the homogeneous waveguide in the acoustic case, we have started a study of the sampling methods in the time domain. This will be the subject of Anne-Claire Egloffe's post-doc. The aim is to use the modal formulation of the sampling methods at all frequencies and recompose the best possible image of the defect.

6.6.2. The exterior approach to retrieve obstacles

Participant: Laurent Bourgeois.

This theme is a collaboration with Jérémi Dardé from IMT (Toulouse).
We have adapted the exterior approach developped for the Laplace equation to the Stokes system. The aim is to find a fixed Dirichlet obstacle in a fluid which is governed by the Stokes system with the help of boundary measurements. The exterior approach consists in defining a decreasing sequence of domains that converge in some sense to the obstacle. More precisely, such iterative approach is based on a combination of a quasi-reversibility method to update the solution of the ill-posed Cauchy problem outside the obstacle obtained at previous iteration and of a level set method to update the obstacle with the help of the solution obtained at previous iteration. In particular, we have introduced two different mixed formulations of quasi-reversibility for the ill-posed Stokes systems in order to use standard Lagrange finite elements.

6.6.3. Inverse scattering with generalized impedance boundary conditions

Participants: Laurent Bourgeois, Mathieu Chamaillard, Nicolas Chaulet.

This work is a collaboration between POEMS and DEFI projects (more precisely Houssem Haddar) and constitutes the subject of the PhD thesis of N. Chaulet, which was defended on the 27/11/2012. We are concerned with the identification of some obstacle and some Generalized Impedance Boundary Conditions (GIBC) on the boundary of such obstacle from far field measurements generated by the scattering of harmonic incident waves. The GIBCs are approximate models for thin coatings, corrugated surfaces, rough surfaces or imperfectly conducting media.

During this last year, we complemented our previous work in two directions. First, we justified the use of the Factorization method to solve the inverse obstacle problem in the presence of GIBCs. This method gives a uniqueness proof as well as a fast algorithm to reconstruct the obstacle from the knowledge of the far field produced by incident plane waves for all the directions of incidence at a given frequency. We also provided some numerical reconstructions of obstacles for several impedance operators.

Meanwhile, we studied the application of non linear optimization techniques to solve the inverse problem for the $3D$ Maxwell's equations. The main advantage of this type of method is that they can be applied with much less data than the Factorization method. Nevertheless, we had to compute the partial derivatives of the electromagnetic field with respect to the parameters we want to reconstruct. In our case, these parameters are the coefficients that define the impedance operator and the shape of the obstacle. We characterized these derivatives in the case where the GIBC is defined by a second order surface operator. The applicability of such methods has been illustrated by some numerical experiments in dimension 3 in which we reconstructed the shape of the scatterer as well as the coefficients that characterize the impedance operator. As demonstrated in the two dimensional case, we think that the GIBCs could be efficiently used to identify the shape of coated objects as well as the parameters of the coating in the $3D$ Maxwell case.

6.6.4. Linear sampling methods in the time domain

Participant: Simon Marmorat.

This work is developed in collaboration with H. Haddar (DEFI, Inria Saclay) and A. Lechleiter (Bremen University). We are concerned with the inverse problem of reconstructing obstacles from the knowledge of scattered acoustic waves in the time domain. We tackle this problem using a linear sampling method that directly acts on time domain data: this imaging technique yields a picture of the scatterer by solving a linear operator equation involving the measured data for many right-hand sides given by singular solutions to the wave equation. We have illustrated the method on numerical examples and have shown a good behaviour with respect to aperture (the quality of reconstruction is better than in the frequency case in the case of limited aperture) and the ability of simultaneously reconstructing obstacles with different boundary conditions among the Dirichlet, Neumann and Robin-Fourier ones.

6.6.5. Space-time focusing on unknown scatterers

Participants: Maxence Cassier, Patrick Joly, Christophe Hazard.

This topic concerns the studies started two years ago about time-reversal in the context of Maxence Cassier's thesis. The main question is to generate a time-dependent wave that focuses on one given scatterer not only in space, but also in time. Our recent works concern two items. On one hand, we have proposed a way to construct such a focusing wave which does not require an a priori knowledge of the location of the obstacle. This wave is represented by a suitable superposition of the eigenvectors of the so-called time-reversal operator in the frequency domain. Numerical results show the focusing properties of such a wave. On the other hand, we try to understand how to translate the physical idea of ìfocusingî into mathematical terms. We proposed and and implemented energy criterion which can be used in numerical experiments in order to evaluate the quality of the focus.

6.6.6. Asymptotic analysis of the interior transmission eigenvalues related to coated obstacles

Participant: Nicolas Chaulet.

This work is a collaboration with Fioralba Cakoni from the University of Delaware (USA) and Houssem Haddar from the DEFI project. The interior transmission eigenvalues play an important role in the area of inverse scattering problems. These eigenvalues can actually be determined by multi-static far field data. Thus, they could be used for non destructive testing. We focused on the case where the obstacle is a perfectly conducting body coated by some thin dielectric material. We derived and justified the asymptotic expansion of the first interior transmission eigenvalue with respect to the thickness of the coating for the TM electromagnetic polarization. This expansion provided interesting qualitative information about the behavior of these eigenvalues and also gave an explicit formula to compute the thickness of the coating.

6.6.7. Interior transmission problem

Participants: Anne-Sophie Bonnet-Ben Dhia, Lucas Chesnel, Jérémi Firozaly.

This work is a collaboration with F. Cakoni from the University of Delaware (U.S.) and H. Haddar from the DEFI project at Inria Saclay. The interior transmission problem plays an important role in the inverse scattering theory for inhomogeneous media. In particular, it arises when one is interested in the reconstruction of an inclusion embedded in a background medium from multi-static measurements of diffracted fields at a given frequency. Physically, it is important to prove that, for a given frequency, there are no waves which do not scatter. Mathematically, this last property boils down to state that the frequency is not a transmission eigenvalue, that is, an eigenvalue of the interior transmission problem. An important issue is to prove that transmission eigenvalues form at most a discrete set with infinity as the only accumulation point. This is not straightforward because the operator associated with this problem exhibits a sign changing in its principal part and its study is not standard. Using the T-coercivity approach, we proved the discreteness under relatively weak assumptions both for the scalar and Maxwell cases. In particular, the simple technique we proposed allows to treat cases, which were not covered by existing methods, where the difference between the inclusion index and the background index changes sign. Now, we are trying to understand the fundamental links which exist between this problem and the transmission problem between a positive and a negative material. In some configurations, the study of the interior transmission problems leads to consider the operator $\Delta(\sigma\Delta\cdot) : H_0^2(\Omega) \to H^{-2}(\Omega)$ where Ω is the domain and σ is a coefficient which changes sign on Ω. During the internship of Jérémy Firozaly, we proved that this operator exhibits properties very different from the operator $\mathrm{div}(\sigma\nabla\cdot) : H_0^1(\Omega) \to H^{-1}(\Omega)$.

6.6.8. Flaw identification using elastodynamic topological derivative

Participants: Marc Bonnet, Rémi Cornaggia.

In collaboration with Cédric Bellis (Columbia Univ. USA), Bojan Guzina (Univ. of Minnesota, USA).
The concept of topological derivative (TD) quantifies the perturbation induced to a given cost functional by the nucleation of an infinitesimal flaw in a reference defect-free body, and may serve as a flaw indicator function. In this work, the TD is derived for three-dimensional crack identification exploiting over-determined transient elastodynamic boundary data. This entails in particular the derivation of the relevant polarization tensor, here given for infinitesimal trial cracks in homogeneous or bi-material elastic bodies. Simple and efficient adjoint-state based formulations are used for computational efficiency, allowing to compute the TD field for arbitrarily shaped elastic solids. The latter is then used as an indicator function for the spatial location of the sought crack(s). Current investigations focus on justifying the heuristic underpinning TD-based identification, which consists in deeming regions where the TD is most negative as the likeliest locations of actual flaws and on formulating higher-order topological expansions in the elastodynamic case.

6.6.9. Topological derivative in anisotropic elasticity

Participant: Marc Bonnet.

In collaboration with Gabriel Delgado (CMAP, Ecole Polytechnique).
Following up on previous work on the topological derivative (TD) of displacement-based cost functionals in anisotropic elasticity, a TD formula has been derived for general cost functionals that involve strains (or displacement gradients) rather than displacements. The small-inclusion asymptotics of such cost functionals are quite different than in the previous case, due to the fact that the strain perturbation inside an elastic inclusion remains finite no matter how small the inclusion size. Cost functionals of practical interest having this format include von Mises equivalent stress (often used in plasticity or failure criteria) and energy-norm error functionals for coefficient-identification inverse problems.

6.6.10. Energy functionals for elastic medium reconstruction using transient data

Participant: Marc Bonnet.

In collaboration with Wilkins Aquino (Cornell Univ., USA).

Energy-based misfit cost functionals, known in mechanics as error in constitutive relation (ECR) functionals, are known since a long time to be well suited to (electrostatic, elastic,...) medium reconstruction. In this ongoing work, a transient elastodynamic version of this methodology is developed, with emphasis on its applicability to large time-domain finite element modeling of the forward problem. The formulation involves coupled transient forward and adjoint solutions, a fact which greatly hinders large-scale computations. A computational approach combining an iterative treatment of the coupled problem and the adjoint to the discrete Newmark time-stepping scheme is found to perform well on large FE models, making the time-domain ECR functional a worthwhile tool for medium identification.

6.7. Other topics

6.7.1. Fast non-overlapping Schwarz domain decomposition methods for the neutron diffusion equation

Participant: Patrick Ciarlet.

A collaboration with Erell Jamelot (CEA Saclay/DEN).
Investigating numerically the steady state of a nuclear core reactor can be very expensive, in terms of memory storage and computational time. In order to address both requirements, one can use a domain decomposition method, which is then implemented on a parallel computer.
We model the problem using a mixed approach, which involves a scalar flux and a vector current. The equivalent variational formulation is then discretized with the help of Raviart-Thomas-Nédélec finite elements. The domain decomposition method is based on the Schwarz iterative algorithm with Robin interface conditions to handle communications. This method is analyzed from the continuous to the discrete point of views: well-posedness, convergence of the finite element method, optimality of the parameter appearing in the Robin interface condition and algorithms. Numerical experiments carried out on realistic 3D configurations using the APOLLO3©code (of CEA/DEN) show the parallel efficiency of the algorithm.

REALOPT Project-Team

6. New Results

6.1. Theoretical and Methodological Developments

Participants: Andrew Miller, Arnaud Pêcher, Pierre Pesneau, Ruslan Sadykov, Gautier Stauffer, François Vanderbeck.

We made progress in the development of theory and algorithms in the area of "Reformulation and Decomposition Approaches for MIP", "Mixed Integer Nonlinear Programming", and "Polyhedral Combinatorics and Graph Theory".

6.1.1. Column Generation for Extended Formulations

Working in an extended variable space allows one to develop tight reformulations for mixed integer programs. However, the size of the extended formulation grows rapidly too large for a direct treatment by a MIP-solver. Then, one can work with inner approximations defined and improved by generating dynamically variables and constraints. The alternative considered in [21] is an inner approximation obtained by generating dynamically the variables of the extended formulation. It assumes that the extended formulation stems from a decomposition principle. Then one can implement column generation for the extended formulation using Dantzig-Wolfe decomposition paradigm. Pricing subproblem solutions are expressed in the variables of the extended formulation and added to the current restricted version of the extended formulation along with the subproblem constraints that are active for the subproblem solutions.

Our paper [21] revisits the column-and-row generation approach, which is viewed herein as a generalization of standard column generation, the latter being based on a specific subproblem extended formulation. This generic view not only highlights the scope of applicability of the method, but it also leads to a more general termination condition than the traditional reduced cost criteria and to theoretically stronger dual bounds. We highlight a key benefit of the latter: lifting pricing problem solutions in the space of the extended formulation permits their recombination into new subproblem solutions and results in faster convergence.

The interest of the approach is evaluated numerically on machine scheduling, bin packing, generalized assignment, and multi-echelon lot-sizing problems. We compare a direct handling of the extended formulation, a standard column generation approach, and the "column-and-row generation" procedure. The results illustrate the stabilization effect resulting from column disaggregation and recombinations that is shown to have a cumulative effect when used in combination with a standard stabilization technique.

6.1.2. Primal Heuristics for Branch-and-Price

Primal heuristics have become an essential component in mixed integer programming (MIP). Generic heuristic paradigms of the literature remain to be extended to the context of a column generation solution approach. Our goal is to derive black-box primal heuristics for use in Branch-and-Price approaches. This requires extending primal heuristic paradigms to the context of dynamic generation of the variables of the model. We highlight an important fact: such generic tools typically performs better than problem specific meta-heuristics, in terms of solution quality and computing times. Based on our application specific experience with these techniques [55], [57], [72], [73], and on a review of generic classes of column generation based primal heuristics, in [49], we are developing a full blown review of such techniques, completed with new methods and an extensive numerical study. This research is being carried on in collaboration with the members of the associated team project, SAMBA [27] [30].

As a Dantzig-Wolfe reformulation is typically tighter than the original compact formulation, techniques based on rounding its linear programming solution have better chance to yield good primal solutions. The aggregated information built into the column definition and the price coordination mechanism provide a global view at the solution space that may be lacking in somewhat more "myopic" approaches based on compact formulations. However, the dynamic generation of variables requires specific adaptation of heuristic paradigms. Our contribution [30] lies in proposing simple strategies to get around these technical issues. We initially concentrate on "diving" methods and consider their combination with "sub-MIPing", relaxation induced neighborhood search, truncated backtracking using a Limited Discrepancy Search. These add-ons serves as local-search or diversification/intensification mechanisms. The methods are numerically tested on standard models such as Cutting Stock, Vertex Coloring, Generalized Assignment, Lot-Sizing, and Vehicle Routing problems. We further extend this research by combining the "diving" method mentioned above with the "feasibility pump" approach [27]. We show how this combination can be implemented in a context of dynamically defined variables, and we report on numerically testing "feasibility pump" for cutting stock and generalized assignment problems.

6.1.3. Stabilization techniques for column generation

Within the SAMBA project, we are collaboratively studying techniques to accelerate the convergence of column generation algorithms [25]. This techniques exploit Lagrangian duality theory. By revisiting all the alternative approaches to solving the Lagrangian dual, we identify suitable combinations of paradigms.

We also bridge the gap with techniques used in the dual framework of cut generation that have their unexploited counterpart for column generation [32], [29]. Cutting plane algorithmic strategies translate into stabilization procedures for column generation. We establish the link between the in-out separation procedure and dual price smoothing techniques for column generation. In this framework, we develop generic convergence proofs and effective smoothing auto-regulating strategies that avoids the need for parameter tuning. We further improve performance of such stabilization by hybridization with an ascent method. This work might inspire novel cut separation strategies.

6.1.4. Stable sets in claw-free graphs

A *stable set* is a set of pairwise non adjacent vertices in a graph and a graph is *claw-free* when no vertex contains a stable set of size three in its neighborhood. Given weights on the vertices, the stable set problem (a NP-hard problem in general) consists in selecting a set of pairwise non adjacent vertices maximizing the sum of the selected weights. The stable set problem in claw-free graphs is a fundamental generalization of the classic matching problem that was shown to be polynomial by Minty in 1980 (G. Minty. *On maximal independent sets of vertices in claw-free graphs*. J. Combinatorial Theory B, 28:284-304 (1980)). However, in contrast with matching, the polyhedral structure (i.e. the integer hull of all stable sets in a claw-free graph) is not very well understood and thus providing a 'decent' linear description of this polytope has thus been a major open problem in our field.

We proposed a new algorithm to find a maximum weighted stable set in a claw-free graph [38] whose complexity is now drastically better than the original algorithm by Minty (n^3 versus n^6, where n is the number of vertices). We also provided a description of the polyhedra in an extended space (i.e. using additional artificial variables) and an *efficient procedure* to separate over the polytope in polynomial-time [26]. Beside those main contributions, we published another papers on the strongly minimal facets of the polytope.

6.1.5. The Circular-Chromatic number

Another central contribution of our team concerns the chromatic number of a graph (the minimum number of independent stable sets needed to cover the graph). We proved that the chromatic number and the clique number of some superclasses of perfect graphs is computable in polynomial time [17].

We investigated the circular-chromatic number. It is a well-studied refinement of the chromatic number of a graph (designed for problems with periodic solutions): the chromatic number of a graph is the integer ceiling of its circular-chromatic number. Xuding Zhu noticed in 2000 that circular cliques are the relevant circular counterpart of cliques, with respect to the circular chromatic number, thereby introducing circular-perfect graphs, a super-class of perfect graphs.

We proved that the clique and chromatic numbers of circular-perfect graphs is computable in polynomial time [16], thereby extending Grötschel, Lovász and Schrijver's result to the whole family of circular-perfect graphs. We gave closed formulas for the Lovász Theta number of circular-cliques (previously, closed formulas were known for circular-cliques with clique number at most 3 only), and derived from them that the circular-chromatic number of circular-perfect graphs is computable in polynomial time [24].

6.2. Model Specific Developments and Applications

Participants: Andrew Miller, Arnaud Pêcher, Pierre Pesneau, Ruslan Sadykov, Gautier Stauffer, François Vanderbeck.

The models on which we made progress can be partitioned in three areas: "Packing and Covering Problems", "Network Design and Routing", and "Planning, Scheduling, and Logistic Problems".

6.2.1. Bin-Packing with Conflicts

The bin-packing problem consists in finding the minimum number of bin of fixed size one needs to pack a set of items of different sizes. We studied a generalization of this problem where items can be in conflicts and thus cannot be put together in the same bin. We show in [20] that the instances of the literature with 120 to 1000 items can be solved to optimality with a generic Branch-and-Price algorithm, such as our prototype BaPCod, within competitive computing time. Moreover, we solved to optimality all the 37 open instances. The approach involves generic primal heuristics, generic branching, but a specific pricing procedure.

6.2.2. Using graph theory for solving orthogonal knapsack problems

We investigated the orthogonal knapsack problem, with the help of graph theory. The multi-dimensional orthogonal packing problem (OPP) is defined as follows: given a set of items with rectangular shapes, the problem is to decide whether there is a non-overlapping packing of these items in a rectangular bin. The rotation of items is not allowed. A powerful characterization of packing configurations by means of interval graphs was introduced by Fekete and Schepers using an efficient representation of all geometrically symmetric solutions by a so called *packing class* involving one *interval graph* (whose complement admits a transitive orientation: each such orientation of the edges corresponds to a specific placement of the forms) for each dimension. Though Fekete & Schepers' framework is very efficient, we have however identified several weaknesses in their algorithms: the most obvious one is that they do not take advantage of the different possibilities to represent interval graphs.

In [12], [11], we give two new algorithms: the first one is based upon matrices with consecutive ones on each row as data structures and the second one uses so-called MPQ-trees, which were introduced by Korte and Mohring to recognize interval graphs. These two new algorithms are very efficient, as they outperform Fekete and Schepers' on most standard benchmarks.

6.2.3. Inventory routing and logistics problems

Inventory routing problems combine the optimization of product deliveries (or pickups) with inventory control at customer sites. in [13], we considered the planning of single product pickups over time: each site accumulates stock at a deterministic rate; the stock is emptied on each visit. Our objective is to minimize a surrogate measure of routing cost while achieving some form of regional clustering by partitioning the sites between the vehicles. The fleet size is given but can potentially be reduced. Planning consists in assigning customers to vehicles in each time period, but the routing, i.e., the actual sequence in which vehicles visit customers, is considered as an "operational" decision. We developed a truncated branch-and-price algorithm. This exact optimization approach is combined with rounding and local search heuristics to yield both primal solutions and dual bounds that allow us to estimate the deviation from optimality of our solution. We were confronted with the issue of symmetry in time that naturally arises in building a cyclic schedule (cyclic permutations along the time axis define alternative solutions). Central to our approach is a state-space relaxation idea that allows us to avoid this drawback: the symmetry in time is eliminated by modeling an average behavior. Our algorithm provides solutions with reasonable deviation from optimality for large scale problems (260 customer sites, 60 time periods, 10 vehicles) coming from industry. The subproblem is interesting in its own right: it is a multiple-class integer knapsack problem with setups. Items are partitioned into classes whose use implies a setup cost and associated capacity consumption.

6.2.4. Scheduling

Cross docking terminals allow companies to reduce storage and transportation costs in a supply chain. At these terminals, products of different types from incoming trucks are unloaded, sorted, and loaded to outgoing trucks for delivery. In [19], we focus on the operational activities at a cross docking terminal with two doors: one for incoming trucks and another one for outgoing trucks. We consider the truck scheduling problem with the objective to minimize the storage usage during the product transfer inside the terminal. Our interest in this problem is mainly theoretical. We show that it is NP-hard in the strong sense even if there are only two product types. For a special case with fixed subsequences of incoming and outgoing trucks, we propose a dynamic programming algorithm, which is the first polynomial algorithm for this case. The results of numerical tests of the algorithm on randomly generated instances are also presented.

In [18], we consider the scheduling jobs in parallel, i.e., jobs can be executed on more than one processor at the same time. With the emergence of new production, communication and parallel computing system, the usual scheduling requirement that a job is executed only on one processor has become, in many cases, obsolete and unfounded. In this work, we consider the NP-hard problem of scheduling malleable jobs to minimize the total weighted completion time (or mean weighted flow time). For this problem, we introduce the class of "ascending" schedules in which, for each job, the number of machines assigned to it cannot decrease over time while this job is being processed. We prove that, under a natural assumption on the processing time functions of jobs, the set of ascending schedules is dominant for the problem. This result can be used to reduce the search space while looking for an optimal solution.

Currently, we are working on a scheduling application at a port. For this application, an equipment routing task scheduling problem [28] has been formulated, where a set of tasks needs to be performed. Tasks require equipment of different types. A particularity of the problem is that an equipment needs to be moved to the actual locations of tasks which use this equipment. So, there are both scheduling and routing decisions are to be taken simultaneously.

6.2.5. One warehouse multi-retailer problem

The *One-Warehouse Multi-retailer problem (OWMR)* is a very important NP-hard inventory control problem arising in the distribution of goods when one central warehouse is supplying a set of final retailers facing demand from customers. In [22], we provide a simple and fast 2-approximation algorithm for this problem (i.e. an algorithm ensuring a deviation by a factor at most two from the optimal solution). This result is both important in practice and in theory as it allows to approximate large real-world instances of the problem (we implemented this algorithm at IBM and it is within 10% of optimality in practice) and the techniques we developed appear to apply to more general settings. We are extending our results to other inventory control problems.

6.3. Software prototypes, Generic Developments and Specific Tools

Participants: Romain LeGuay, Pierre Pesneau, Ruslan Sadykov, François Vanderbeck.

6.3.1. BaPCod - a generic branch-and-price code

The development of the prototype software platform is supported by our junior engineer, Romain Leguay. He developed a new interface with the underlying MIP solver allowing multiple solvers to be called in the same run. He then re-organized the svn depository and a web distribution platform in view of the increasing number if users to whom Romain offers precious support. Romain has then redesigned parts of the code in the perspective of its parallelization and contributed to designing a pseudo modeling language for a friendly user interface. The emphasis is currently on enhancing the code performance in particular through rapid access data structure. Romain also participates to the setting up of stabilization and preprocessing algorithms.

The software platform BaPCod is continuously improved to include all the methodological features that arise from our research, in particular in our collaborative project with Brazil: SAMBA. BaPCod serves there as a proof-of-concept code and is useful for the transfer of knowledge between the parties, including the company GAPSO (a Brazilian spin-up launched by these academics).

REGULARITY Project-Team

6. New Results

6.1. A multifractional Hull and White model

Participants: Joachim Lebovits, Jacques Lévy Véhel.

In collaboration with Sylvain Corlay (Paris 6 University).

We have considered the following model, which is an extension of the fractional Hull and White model proposed in [55]: under the risk-neutral measure, the forward price of a risky asset is the solution of the S.D.E.

$$\begin{aligned} &dF_t = F_t \sigma_t dW_t, \\ &d\ln(\sigma_t) = \theta\left(\mu - \ln(\sigma_t)\right) dt + \gamma_h d^{\diamond} B_t^h + \gamma_\sigma dW_t^\sigma, \quad \sigma_0 > 0, \theta > 0, \end{aligned}$$

where B_t^h is a multifractional Brownian motion with regularity function h, and W_t, W_t^σ are standard Brownian motions. This SDE is interpgreted in the Wick-Itô sense.

Using functional quantization techniques, it is possible to compute numerically implied forward start volatilities for this model. Using an adequate h function estimated from SP500 data, we have shown that this model is able to reproduce to some extent the volatility surface observed on the market [34].

6.2. Markov characterization of the set-indexed Lévy process

Participant: Erick Herbin.

In collaboration with Prof. Ely Merzbach (Bar Ilan university, Israel).

In [21], the class of set-indexed Lévy processes is considered using the stationarity property defined for the set-indexed fractional Brownian motion in [20]. The general framework of Ivanoff-Merzbach allows to consider standard properties of stochastic processes (e.g. martingale and Markov properties) in the set-indexed context. Processes are indexed by a collection $\mathcal{A}$ of compact subsets of a metric space $\mathcal{T}$ equipped with a Radon measure m, which satisfies several stability conditions. Each process $\{X_U;\ U \in \mathcal{A}\}$ is assumed to admit an increment process $\{\Delta X_C;\ C \in \mathcal{C}\}$ defined as an additive extension of X to the collections $\mathcal{C}_0 = \{U \setminus V;\ U, V \in \mathcal{A}\}$ and

$$\mathcal{C} = \left\{ U \setminus \bigcup_{1 \leq i \leq n} V_i;\ n \in \mathbf{N}; U, V_1, \cdots, V_n \in \mathcal{A} \right\}.$$

A set-indexed process $X = \{X_U;\ U \in \mathcal{A}\}$ is called a *set-indexed Lévy process* if the following conditions hold

1. $X_{\varnothing'} = 0$ almost surely, where $\varnothing' = \bigcap_{U \in \mathcal{A}} U$.
2. the increments of X are independent: for all pairwise disjoint $C_1, \cdots, C_n$ in $\mathcal{C}$, the random variables $\Delta X_{C_1}, \cdots, \Delta X_{C_n}$ are independent.
3. X has m-stationary $\mathcal{C}_0$-increments, i.e. for all integer n, all $V \in \mathcal{A}$ and for all increasing sequences $(U_i)_i$ and $(A_i)_i$ in $\mathcal{A}$, we have

$$[\forall i,\ m(U_i \setminus V) = m(A_i)] \Rightarrow (\Delta X_{U_1 \setminus V}, \cdots, \Delta X_{U_n \setminus V}) \stackrel{(d)}{=} (\Delta X_{A_1}, \cdots, \Delta X_{A_n})$$

4. X is continuous in probability: if $(U_n)_{n \in \mathbf{N}}$ is a sequence in $\mathcal{A}$ such that

$$\overline{\bigcup_n \bigcap_{k\geq n} U_k} = \bigcap_n \overline{\bigcup_{k\geq n} U_k} = A \in \mathcal{A}$$

then

$$\lim_{n\to\infty} P\left\{|X_{U_n} - X_A| > \epsilon\right\} = 0$$

On the contrary to previous works of Adler and Feigin (1984) on one hand, and Bass and Pyke (1984) one the other hand, the increment stationarity property allows to obtain explicit expressions for the finite-dimensional distributions of a set-indexed Lévy process. From these, we obtained a complete characterization in terms of Markov properties.

Among the various definitions for Markov property of a SI process, we considered the $\mathcal{Q}$-Markov property. A collection $\mathcal{Q}$ of functions

$$\begin{aligned} \mathbf{R} \times \mathcal{B}(\mathbf{R}) &\to \mathbf{R}_+ \\ (x, B) &\mapsto Q_{U,V}(x, B) \end{aligned}$$

where $U, V \in \mathcal{A}(u)$ are s.t. $U \subseteq V$, is called a *transition system* if the following conditions are satisfied:

1. $Q_{U,V}(\bullet, B)$ is a random variable for all $B \in \mathcal{B}(\mathbf{R})$.
2. $Q_{U,V}(x, \bullet)$ is a probability measure for all $x \in \mathbf{R}$.
3. For all $U \in \mathcal{A}(u)$, $x \in \mathbf{R}$ and $B \in \mathcal{B}(\mathbf{R})$, $Q_{U,U}(x, B) = \delta_x(B)$.
4. For all $U \subseteq V \subseteq W \in \mathcal{A}(u)$,

$$\int_{\mathbf{R}} Q_{U,V}(x, dy) Q_{V,W}(y, B) = Q_{U,W}(x, B).$$

A transition system $\mathcal{Q}$ is said

- *spatially homogeneous* if for all $U \subset V$,

$$\forall x \in \mathbf{R}, \forall B \in \mathcal{B}(\mathbf{R}), \quad Q_{U,V}(x, B) = Q_{U,V}(0, B - x);$$

- *m-homogeneous* if $Q_{U,V}$ only depends on $m(V \setminus U)$,
  i.e. $\forall U, V, U', V' \in \mathcal{A}(u)$ such that $U \subset V$ and $U' \subset V'$,

$$m(V \setminus U) = m(V' \setminus U') \Rightarrow Q_{U,V} = Q_{U',V'}.$$

A set-indexed process $X := \{X_U;\ U \in \mathcal{A}\}$ is called *$\mathcal{Q}$-Markov* if $\forall U, V \in \mathcal{A}(u), U \subseteq V$

$$\forall B \in \mathcal{B}(\mathbf{R}), \quad P[\Delta X_V \in \Gamma \mid \mathcal{F}_U] = Q_{U,V}(\Delta X_U; \Gamma),$$

where $(\mathcal{F}_U)_{U\in\mathcal{A}(u)}$ is the minimal filtration of the process X.

Balan-Ivanoff (2002) proved that any SI process with independent increments is a $\mathcal{Q}$-Markov process with a spatially homogeneous transition system. The following result proved in [21] shows that the converse is true.

Theorem *Let $X = \{X_U;\ U \in \mathcal{A}\}$ be a set-indexed process with definite increments. The two following assertions are equivalent:*

1. *X is a $\mathcal{Q}$-Markov process with a spatially homogeneous transition system $\mathcal{Q}$;*
2. *X has independent increments.*

This result is strengthened in the following characterization of set-indexed Lévy processes as Markov processes with homogeneous transition systems.
Theorem *Let $X = \{X_U;\ U \in \mathcal{A}\}$ be a set-indexed process with definite increments and satisfying the stochastic continuity property.*

The two following assertions are equivalent:

1. *X is a set-indexed Lévy process ;*
2. *X is a $\mathcal{Q}$-Markov process such that $X_\varnothing = 0$ and the transition system $\mathcal{Q}$ is spatially homogeneous and m-homogeneous.*

Consequently, if $\mathcal{Q}$ is a transition system which is both spatially homogeneous and m-homogeneous, then there exists a set-indexed process X which is a $\mathcal{Q}$-Markov process.

6.3. Local Hölder regularity of Set-Indexed processes

Participants: Erick Herbin, Alexandre Richard.

In the set-indexed framework of Ivanoff and Merzbach ([62]), stochastic processes can be indexed not only by $\mathbf{R}$ but by a collection $\mathcal{A}$ of subsets of a measure and metric space $(\mathcal{T}, d, m)$, with some assumptions on $\mathcal{A}$. In we introduce and study some assumptions (A_1) and (A_2) on the metric indexing collection $(\mathcal{A}, d_{\mathcal{A}})$ in order to obtain a Kolmogorov criterion for continuous modifications of SI stochastic processes. Under this assumption, the collection is totally bounded and a set-indexed process with good incremental moments will have a modification whose sample paths are almost surely Hölder continuous, for the distance $d_{\mathcal{A}}$. Once this condition is established, we investigate the definition of Hölder coefficients for SI processes. We shall denote $\widetilde{\alpha}_X(t)$ and $\alpha_X(t)$ for the local and pointwise Hölder exponents of X at t, and $\widetilde{\alpha}_X(t)$ and $\alpha_X(t)$ for their deterministic counterpart in case X is Gaussian.

In [18], a set-indexed extension for fractional Brownian motion has been defined and studied. A mean-zero Gaussian process $\mathbf{B}^H = \left\{\mathbf{B}_U^H, U \in \mathcal{A}\right\}$ is called a *set-indexed fractional Brownian motion (SIfBm for short)* on $(\mathcal{T}, \mathcal{A}, m)$ if

$$\forall U, V \in \mathcal{A}, \quad \mathbf{E}\left[\mathbf{B}_U^H \mathbf{B}_V^H\right] = \frac{1}{2}\left[m(U)^{2H} + m(V)^{2H} - m(U \bigtriangleup V)^{2H}\right], \tag{11}$$

where $H \in (0, 1/2]$ is the index of self-similarity of the process.

In [12], $\widetilde{\alpha}_X$ and $\widetilde{\alpha}_X$ have been determined for the particular case of an SIfBm indexed by the collection $\{[0,t];\ t \in \mathbf{R}_+^N\} \cup \{\varnothing\}$, called the *multiparameter fractional Brownian motion*. If X denotes the $\mathbf{R}_+^N$-indexed process defined by $X_t = \mathbf{B}_{[0,t]}^H$ for all $t \in \mathbf{R}_+^N$, it is proved that for all $t_0 \in \mathbf{R}_+^N$, $\widetilde{\alpha}_X(t_0) = H$ and with probability one, for all $t_0 \in \mathbf{R}_+^N$, $\widetilde{\alpha}_X(t_0) = H$. A theorem of allows one to extend these results to SIfBm indexed by a more general class than the sole collection of rectangles of $\mathbf{R}_+^N$.

Theorem 0.1 Let $\mathbf{B}^H$ be a set-indexed fractional Brownian motion on $(\mathcal{T}, \mathcal{A}, m)$, $H \in (0, 1/2]$. Assume that the subclasses $(\mathcal{A}_n)_{n\in\mathbf{N}}$ satisfy Assumption (A_1).
Then, the local and pointwise Hölder exponents of $\mathbf{B}^H$ at any $U_0 \in \mathcal{A}$, defined with respect to the distance d_m or any equivalent distance, satisfy

$$\mathbf{P}\left(\forall U_0 \in \mathcal{A},\ \widetilde{\alpha}_{\mathbf{B}^H}(U_0) = H\right) = 1$$

and if Assumption (A_2) holds,

$$\mathbf{P}\left(\forall U_0 \in \mathcal{A},\ \alpha_{\mathbf{B}^H}(U_0) = H\right) = 1.$$

Consequently, since the collection $\mathcal{A}$ of rectangles of $\mathbf{R}_+^N$ with m the Lebesgue measure satisfies (A_1) and (A_2), we obtained a new result on a classical multiparameter process: the multiparameter fractional Brownian motion $\mathbf{B}^H$ satisfy, for $T \in \mathbf{R}_+^N$:

$$\mathbf{P}\left(\forall t \in [0,T],\ \alpha_{\mathbf{B}^H}([0,t]) = \widetilde{\alpha}_{\mathbf{B}^H}([0,t]) = H\right) = 1.$$

6.4. Separability of Set-Indexed Processes

Participant: Alexandre Richard.

A classical result states that any (multiparameter) stochastic process has a separable modification, thus ensuring the measurability property of the sample paths. We extend this result to set-indexed processes.
Let $(T, \mathcal{O})$ be a topological space. We assume that this space is *second-countable*, ie there exists a countable subset $\widetilde{\mathcal{O}} \subseteq \mathcal{O}$ such that any open set of $\mathcal{O}$ can be expressed as a union of elements of $\widetilde{\mathcal{O}}$.

A process $\{X_t, t \in T\}$ is *separable* if there exists an at most countable set $S \subset T$ and a null set Λ such that for all closed sets $F \subset \mathbf{R}$ and all open set $O \in \mathcal{O}$,

$$\{\omega : X_s(\omega) \in F forall s \in O \cap S\} \setminus \{\omega : X_s(\omega) \in F forall s \in O\} \subset \Lambda.$$

This definition is different of the one found in [57], where the space is "linear", in that this author considers the previous equation only when O is an interval. It happens that this notion needs not be defined in a general topological space. However when restricted to a vector space, our definition implies the previous one.

Theorem 0.2 (Doob's separability theorem) Any T-indexed stochastic process $X = \{X_t;\ t \in T\}$ has a separable modification.

If T is an indexing collection in the sense of [62], the topology induced by the distance d_T has to be second-countable. This happens for instance when (T, d_T) is totally bounded, which is the case in

6.5. An increment type set-indexed Markov property

Participant: Paul Balança.

[1] investigates a new approach for the definition of a set-indexed Markov property, named $\mathcal{C}$-*Markov*. The study is based on Merzbach and Ivanoff's set-indexed formalism, i.e. $\mathcal{A}$ denotes a set-indexed collection and $\mathcal{C}$ the family of increments $C = A \setminus B$, where $A \in \mathcal{A}$ and $B \in \mathcal{A}(u)$ (finite unions of sets from $\mathcal{A}$). Moreover, for any $C = A \setminus B$, $B = \cup_{i=1}^k A_i$, $\mathcal{A}_{\mathbf{C}}$ is defined as the following subset of $\mathcal{A}$:

$$\mathcal{A}_{\mathbf{C}} = \{U \in \mathcal{A}_\ell;\, U \not\subseteq B^\circ\} := \{U_C^1, \cdots, U_C^p\}, \qquad \text{where } p = |\mathcal{A}_{\mathbf{C}}|$$

and $\mathcal{A}_\ell$ corresponds to the semilattice $\{A_1 \cap \cdots \cap A_k, \cdots, A_1 \cap A_2, A_1 \cdots, A_k\} \subset \mathcal{A}$. The notation $\mathbf{X_C}$ refers to a random vector $\mathbf{X_C} = \left(X_{U_C^1}, \cdots, X_{U_C^p}\right)$. Similarly, $\mathbf{x_C}$ is used to denote a vector of variables $(x_{U_C^1}, \cdots, x_{U_C^p})$.

Then, an E-valued set-indexed process $(X_A)_{A \in \mathcal{A}}$ is said to be $\mathcal{C}$-*Markov* with respect to a filtration $(\mathcal{F}_A)_{A \in \mathcal{A}}$ if it is adapted to $(\mathcal{F}_A)_{A \in \mathcal{A}}$ and if it satisfies

$$\mathbb{E}[f(X_A)\,|\,\mathcal{G}^*_C] = \mathbb{E}[f(X_A)\,|\,\mathbf{X_C}] \quad \mathbb{P}\text{-a.s.} \tag{12}$$

for all $C = A \setminus B \in \mathcal{C}$ and any bounded measurable function $f : E \to \mathbf{R}$. The sigma-algebra $\mathcal{G}^*_C$ is usually called the strong history of $(\mathcal{F}_A)_{A\in\mathcal{A}}$ and is defined as $\mathcal{G}^*_C = \bigvee_{A\in\mathcal{A}, A\cap C=\varnothing} \mathcal{F}_A$.

The $\mathcal{C}$-Markov approach has several advantages compared to existing set-indexed Markov literature (mainly $\mathcal{Q}$-Markov described in [48]). It appears to be a natural extension of the classic one-parameter Markov property. In particular, the concept of transition system can easily extended to our formalism: for any $\mathcal{C}$-Markov process X, one can defined $\mathcal{P} = \{P_C(\mathbf{x_C}; \mathrm{d}x_A); C \in \mathcal{C}\}$ as

$$\forall \mathbf{x_C} \in E^{|\mathcal{A}_\mathbf{C}|}, \Gamma \in \mathcal{E}; \quad P_C(\mathbf{x_C}; \Gamma) := \mathbb{P}(X_A \in \Gamma\,|\,\mathbf{X_C} = \mathbf{x_C}).$$

A $\mathcal{C}$-transition system $\mathcal{P}$ happens to satisfy a set-indexed Chapman-Kolmogorov equation,

$$\forall C \in \mathcal{C},\ A' \in \mathcal{A}; \qquad P_C f = P_{C'} P_{C''} f \quad \text{where} \quad C' = C \cap A',\ C'' = C \setminus A' \tag{13}$$

and f is a bounded measurable function.

Similarly to the classic Markovian theory, is is proved in [1] that the initial distribution μ and $\mathcal{P}$ characterize entirely the law of a $\mathcal{C}$-Markov process, and that conversely, for any initial law and any $\mathcal{C}$-transition system, a corresponding canonical set-indexed $\mathcal{C}$-Markov process can be constructed. $\mathcal{C}$-Markov processes enjoy several other properties such as

1. Projections on elementary flows are Markovian;
2. Conditional independence of natural filtrations;
3. Strong Markov property.

The class of set-indexed Lévy processes defined and studied in [21] offers examples of $\mathcal{C}$-Markov processes whose transition probabilities correspond to

$$\forall C = A \setminus B \in \mathcal{C}, \quad \forall \Gamma \in \mathcal{E}; \quad P_C(\mathbf{x_C}; \Gamma) = \mu^{m(C)}(\Gamma - \Delta x_B), \tag{14}$$

where m is a measure on $\mathcal{T}$ and μ the infinitely divisible probability measure that characterizes the Lévy process. We note that the transition system related the $\mathcal{Q}$-Markov property has a different form, even if it is related.

Another non-trivial example of $\mathcal{C}$-Markov process is the set-indexed Ornstein-Uhlenbeck process that has been introduced and studied in [32]. It is a Gaussian Markovian process whose transition densities are given by

$$p_C(\mathbf{x_C}; y) = \frac{1}{\sigma_C\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_C^2}\left(y - e^{-\lambda m(A)}\left[\sum_{i=1}^n (-1)^{\varepsilon_i}\, x_{U_C^i}\, e^{\lambda m(U_C^i)}\right]\right)^2\right], \tag{15}$$

where λ and σ are positive parameters, m is a measure on $\mathcal{T}$ and

$$\sigma_C^2 = \frac{\sigma^2}{2\lambda}\left(1 - e^{-2\lambda m(A)}\left[\sum_{i=1}^n (-1)^{\varepsilon_i} e^{2\lambda m(U_C^i)}\right]\right).$$

In the particular case of multiparameter processes, corresponding to the indexing collection $\mathcal{A} = \{[0,t]; t \in \mathbf{R}_+^N\}$, the $\mathcal{C}$-Markov formalism is related to several existing works. It generalizes the two-parameter $*$-Markov property introduced in [53] and also embraces the multiparameter Markov property investigated recently in [68]. Finally, under some Feller assumption on the transition system, a multiparameter $\mathcal{C}$-Markov process is proved to admit a modification with right-continuous sample paths.

6.6. Fine regularity of Lévy processes

Participant: Paul Balança.

This ongoing work focuses on the fine regularity of one-parameter Lévy processes. The main idea of this study is to use the framework of stochastic 2-microlocal analysis (introduced and developed in [16],[33]) to refine sample paths results obtained in [65].

The latter describes entirely the multifractal spectrum of Lévy processes, i.e. the Hausdorff geometry of level sets $(E_h)_{h\in\mathbf{R}_+}$ of the pointwise exponent. These are usually called the *iso-Hölder sets* of X and are given by

$$E_h = \{t \in \mathbf{R} : \alpha_{X,t} = h\} \quad \text{for every } h \in \mathbf{R}_+ \cup \{+\infty\}.$$

The multifractal spectrum is itself defined as the localized Hausdorff dimension of the previous sets, i.e.

$$d_X(h,V) = \dim_{\mathcal{H}}(E_h \cap V) \quad \text{for every } h \in \mathbf{R}_+ \cup \{+\infty\} \text{ and } V \in \mathcal{O} \text{ (open sets in } \mathbf{R}). \tag{16}$$

[65] states that under a mild assumption on the Lévy measure π, a Lévy process X with no Brownian component almost surely satisfies

$$\forall V \in \mathcal{O}; \quad d_X(h,V) = \begin{cases} \beta h & \text{if } h \in [0,1/\beta]; \\ -\infty & \text{if } h \in (1/\beta,+\infty], \end{cases} \tag{17}$$

where the Blumenthal-Getoor exponent β is given by

$$\beta = \inf\left\{\delta \geq 0 : \int_{\mathbf{R}^d} \left(1 \wedge \|x\|^\delta\right) \pi(\mathrm{d}x) < \infty\right\}. \tag{18}$$

Since classic multifractal analysis focuses on the pointwise exponent, it is natural from our point of view to integrate the 2-microlocal frontier into this description. More precisely, we focus on the dichotomy usual/unusual regularity, corresponding to the sets $(\widetilde{E}_h)_{h\in\mathbf{R}_+}$ and $(\widehat{E}_h)_{h\in\mathbf{R}_+}$:

$$\widetilde{E}_h = \{t \in E_h : \forall s' \in \mathbf{R};\ \sigma_{X,t}(s') = (h+s') \wedge 0\} \quad \text{and} \quad \widehat{E}_h = E_h \setminus \widetilde{E}_h.$$

The collection $(\widetilde{E}_h)_{h\in\mathbf{R}_+}$ represents times at which the 2-microlocal behaviour is rather common (i.e. the slope is equal to one), whereas at points which belong $(\widehat{E}_h)_{h\in\mathbf{R}_+}$, the 2-microlocal frontier has an unusual form.

Then, our main result states that sample paths of a Lévy process X with no Brownian component almost surely satisfy

$$\forall V \in \mathcal{O}; \quad \dim_{\mathcal{H}}(\widetilde{E}_h \cap V) = \begin{cases} \beta h & \text{if } h \in [0,1/\beta]; \\ -\infty & \text{if } h \in (1/\beta,+\infty]. \end{cases} \tag{19}$$

Furthermore, the collection of sets $(\widehat{E}_h)_{h\in\mathbf{R}_+}$ enjoys almost surely

$$\forall V \in \mathcal{O}; \quad \dim_{\mathcal{H}}(\widehat{E}_h \cap V) \leq \begin{cases} 2\beta h - 1 & \text{if } h \in (1/2\beta, 1/\beta)\,; \\ -\infty & \text{if } h \in [0, 1/2\beta] \cup [1/\beta, +\infty]. \end{cases} \tag{20}$$

These results clearly extend those obtained in [65] since we know that the pointwise exponent is completely characterize by the 2-microlocal frontier. Moreover, it also proves that from a Hausdorff dimension point of view, the common regularity is a 2-microlocal frontier with a slope equal to one.

Nevertheless, equation (15) also exhibits some unusual behaviours, corresponding to times $(\widehat{E}_h)_{h\in\mathbf{R}_+}$, that are not captured by the classic multifractal spectrum. The existence of such particular times highly depends on the structure of the Lévy measure, and not only the value of the Blumenthal-Getoor exponent which is therefore not sufficient to characterize entirely the fine regularity. This last aspect of the study illustrates the fact that 2-microlocal analysis is an interesting tool for the study of stochastic processes' regularity since some sample paths' properties can not be captured by common tools such as Hölder exponents.

6.7. A class of self-similar processes with stationary increments in higher order Wiener chaoses.

Participant: Benjamin Arras.

Self similar processes with stationary increments (SSSI processes) have been studied for a long time due to their importance both in theory and in practice. Such processes appear as limits in various renormalisation procedures [69]. In applications, they occur in various fields such as hydrology, biomedicine and image processing. The simplest SSSI processes are simply Brownian motion and, more generally, Lévy stable motions. Apart from these cases, the best known such process is probably fractional Brownian motion (fBm). A construction of SSSI processes that generalizes fBm to higher order Wiener chaoses was proposed in [73]. These processes read

$$\forall t \in \mathbb{R}_+ \quad X_t = \int_{\mathbb{R}^d} h_t^H(x_1, ..., x_d) dB_{x_1} ... dB_{x_d}$$

where h_t^H verifies:

1. $h_t^H \in \widehat{L}^2(\mathbb{R}^d)$,
2. $\forall c > 0, \quad h_{ct}^H(cx_1, ..., cx_d) = c^{H-\frac{d}{2}} h_t^H(x_1, ..., x_d)$,
3. $\forall \rho \geq 0, \quad h_{t+\rho}^H(x_1, ..., x_d) - h_t^H(x_1, ..., x_d) = h_\rho^H(x_1 - t, ..., x_d - t)$.

In [41], we define a class of such processes by the following multiple Wiener-Itô integral representation:

$$X_t^\alpha = \int_{\mathbb{R}^d} \left[||\mathbf{t}^* - \mathbf{x}||_2^{H-\frac{d}{2}} - ||\mathbf{x}||_2^{H-\frac{d}{2}} \right] dB_{x_1} ... dB_{x_d} \tag{21}$$

where $t \in [0,1]$, $\mathbf{t}^* = (t, ..., t)$ and $\alpha = H - 1 + \frac{d}{2}$. When $d = 1$, this is just fBm. In order to study the local regularity of this class of processes as well as the asymptotic behaviour at infinity, we use wavelet's methods. More precisely, following ideas from [46], we obtain the following wavelet-like expansion:
Almost surely,

$$\forall t \in [0,1] \quad X_t^\alpha = \sum_{j\in\mathbb{Z}} \sum_{\mathbf{k}\in\mathbb{Z}^d} \sum_{\epsilon\in E} 2^{-jH} \left[I^{\alpha+1}(\psi^{(\epsilon)})(2^j \mathbf{t}^* - \mathbf{k}) - I^{\alpha+1}(\psi^{(\epsilon)})(-\mathbf{k}) \right] I_d(\psi_{j,\mathbf{k}}^{(\epsilon)}).$$

From this representation, we get several results about this class of processes. Namely:

- There exists a strictly positive random variable A_d of finite moments of any order and a constant, $b_d > 1$, such that:

$$\forall \omega \in \Omega^* \quad \sup_{(s,t)\in[0,1]} \frac{|X_t^\alpha(\omega) - X_s^\alpha(\omega)|}{|t-s|^H (\log(b_d + |t-s|^{-1}))^{\frac{d}{2}}} \le A_d(\omega).$$

- There exists a strictly positive random variable B_d of finite moments of any order and a constant $c_d > 3$, such that:

$$\forall \omega \in \Omega^* \quad \sup_{t\in\mathbb{R}_+} \frac{|X_t^\alpha(\omega)|}{(1+|t|)^H (\log\log(c_d + |t|))^{\frac{d}{2}}} \le B_d(\omega).$$

Using an estimate from [54], we compute the uniform almost sure pointwise Hölder exponent of X^α defined by:

$$\gamma_{X^\alpha}(t) = \sup\{\gamma > 0 : \quad \limsup_{\rho\to 0} \frac{|X_{t+\rho}^\alpha - X_t^\alpha|}{|\rho|^\gamma} < +\infty\}.$$

We get the following result:
Almost surely,

$$\forall t \in (0,1), \ \gamma_{X^\alpha}(t) = H.$$

In the last part of [41], we give general bounds on the Hausdorff dimension of the range and graphs of multidimensional anisotropic SSSI processes defined by multiple Wiener integrals. Let $Y_t^H = \gamma(H,d) I_d(h_t^H)$ where $\gamma(H,d)$ is a normalizing positive constant such that $\mathbb{E}[|Y_1^H|^2] = 1$. Let $\frac{1}{2} < H_1 \le ... \le H_N < 1$. Let $\{\mathbb{Y}_t^H\}$ be the multidimensional process defined by:

$$\{\mathbb{Y}_t^H\} = \{(Y_t^{H_1}, ..., Y_t^{H_N}) : t \in \mathbb{R}_+\}$$

where the coordinates are independent copies of the process Y_t^H. Following classical ideas from [78] and using again the estimate from [54], we obtain:
Almost surely,

$$dim_{\mathcal{H}} R_E(\mathbb{Y}^H) \ge \min\left(N; \frac{dim_{\mathcal{H}} E + \frac{\sum_{j=1}^k (H_k - H_j)}{d}}{H_k}, k = 1, ..., N\right),$$

$$dim_{\mathcal{H}} Gr_E(\mathbb{Y}^H) \ge \min\left(\frac{dim_{\mathcal{H}} E + \frac{\sum_{j=1}^k (H_k - H_j)}{d}}{H_k}, k = 1, ..., N, dim_{\mathcal{H}} E + \sum_{i=1}^N \frac{(1-H_i)}{d}\right).$$

And,

$$dim_{\mathcal{H}} R_E(\mathbb{Y}^H) \leq \min\left(N; \frac{dim_{\mathcal{H}} E + \sum_{j=1}^{k}(H_k - H_j)}{H_k}, k = 1, ..., N\right),$$

$$dim_{\mathcal{H}} Gr_E(\mathbb{Y}^H) \leq \min\left(\frac{dim_{\mathcal{H}} E + \sum_{j=1}^{k}(H_k - H_j)}{H_k}, k = 1, ..., N; dim_{\mathcal{H}} E + \sum_{i=1}^{N}(1 - H_i)\right).$$

where $E \subset \mathbb{R}_+$.

6.8. Economic growth models

Participants: Jacques Lévy Véhel, Lining Liu.

In collaboration with D. La Torre, University of Milan.

We study certain economic growth models where we add a source of randomness to make the evolution equations more realistic. We have studied two particular models:

- An augmented Uzawa-Lucas growth model where technological progress is modelled as the solution of a stochastic differential equation driven by a Lévy or an additive process. This allows for a more faithful description of reality by taking into account discontinuities in the evolution of the level of technology. In details, we consider a closed economy in which there is single good which is produced by combining physical capital $K(t)$ and human capital $H(t)$. The laws of motions of $K(t)$ and $H(t)$ are:

$$\dot{K}(t) = A(t)^{\gamma}[u(t)H(t)]^{\xi} K(t)^{1-\xi-\gamma} - \beta_K K(t) - C(t), \tag{22}$$

$$K(0) = K_0;$$

$$\dot{H}(t) = (\eta(1 - u(t)) - \beta_H) H(t), \tag{23}$$

$$H(0) = H_0,$$

where $A(t)$ is the level of technology, $H(t)$ is the total stock of human capital, $u(t)$ is the proportion to the production of good, $\gamma \in (0,1)$, $\xi \in (0,1)$ and $1 - \xi - \gamma \in (0,1)$ are the shares of income accruing to $A(t)$, $u(t)H(t)$ and $K(t)$, respectively, $\beta_K \in [0,1]$ is the constant rate of depreciation of physical capital, $\beta_H \in [0,1]$ is the rate of depreciation of human capital and $\eta \geq 0$ is the productivity of human capital.

We assume that the level of technology evolves according to the following stochastic differential equation:

$$dA(t) = \mu A(t)dt + \sigma A(t)dW(t) + \delta \int A(t^-)z(\widetilde{N}(dt, dz) - \nu(dt, dz)), \tag{24}$$

where $\mu \in \mathbb{R}$ is the drift rate, $\sigma > 0$ is the volatility, $0 \leq \delta \leq 1$, W is a standard Brownian motion and $\widetilde{N}$ is Poisson random measure with intensity measure ν which satisfies

$$\lim_{s\to t^+}\frac{1}{s-t}\int_t^s\int_{-1}^1 z^2\nu(dz,dx)+\lim_{s\to t^+}\frac{1}{s-t}\int_t^s\int_1^\infty z\nu(dz,dx)<\infty,$$

and

$$\int_0^t\int_{-1}^1 z^2\nu(dz,dx)+\int_0^t\int_1^\infty z\nu(dz,dx)<\infty,$$

for $t>0$.

With a CIES utility function, the optional inter-temporal decision problem can be formulated as

$$\max_{[C,u]}\mathbb{E}\left[\int_0^\infty\frac{C(t)^{1-\phi}-1}{1-\phi}e^{-\rho t}dt\right],\tag{25}$$

where $\rho>0$ is the rate of time preference and $\phi>0$. We denote $V(H,K,A)$ the maximum value function associated with the stochastic optimisation problem. For given t, the maximum expect utility up to time t obtained when applying the stochastic control $[C(t),u(t)]$ is defined by

$$V(H(t),K(t),A(t))=\max_{[C,u]}\mathbb{E}\left[\int_0^t\frac{C(x)^{1-\phi}-1}{1-\phi}e^{-\rho x}dx\right].\tag{26}$$

We have been able to solve this program under some simplifying assumptions. Numerical simulations allow one to assess precisely the effect of (tempered) multistable noise on the model.

- A stochastic demographic jump shocks in a multi-sector growth model with physical and human capital accumulation. This models allows one to take into account sudden changes in population size, due for instance to wars or natural catastrophes. The laws of motions of physical capital $K(t)$ and human capital $H(t)$ are:

$$\dot{K}(t)=AM(t)^{1-\xi-\beta}[u(t)H(t)]^\beta K(t)^\xi-\eta_K K(t)-c(t)M(t),\tag{27}$$

$$\dot{H}(t)=B(1-u(t))H(t)-\eta_H H(t),\tag{28}$$

with initial conditions $K(0)=K_0$ and $H(0)=H_0$, where $M(t)$ is the population size, $H(t)$ is the human capital, $u(t)$ is the share of human capital employed in production, $\beta\in(0,1)$, $\xi\in(0,1)$ and $1-\xi-\beta\in(0,1)$ are the shares accruing to $M(t)$, $u(t)H(t)$ and $K(t)$, respectively, $\eta_K\in[0,1]$ is the constant rate of depreciation of physical capital, $\eta_H\in[0,1]$ is the rate of depreciation of human capital and $A\geq 0$, $B\geq 0$ are the productivities of physical capital and human capital.

We assume that the population size evolves according to the following stochastic differential equation:

$$dM(t)=\mu M(t)dt+\sigma M(t)dW(t)+\delta\int M(t^-)z(\widetilde{N}(dt,dz)-\nu(dt,dz))$$

with initial condition $M(0)=M_0$, where $\mu\in\mathbb{R}$ is the drift rate, $\sigma>0$ is the volatility, $0\leq\delta\leq 1$, W is a standard Brownian motion and $\widetilde{N}$ is Poisson random measure with intensity measure $\nu(dt,dz)$.

Here again, we are able to solve an optimisation program under some simplifying assumptions. This sheds light on the effect of demographic shocks on macroeconomic growth.

SCIPORT Team

6. New Results

6.1. Automatic Differentiation and parallel codes

Participants: Valérie Pascual, Laurent Hascoët, Hubert Alcin, Jean Utke [Argonne National Lab. (Illinois, USA)], Uwe Naumann [RWTH Aachen University (Germany)].

Together with colleagues in Argonne National Lab. and RWTH Aachen, we are studying how AD tools can handle MPI-parallel codes, especially in adjoint mode.

This year, we have presented our strategy [16] to extend Data-Flow analysis to Message-Passing communication. This strategy is specially designed for a program representation like that of TAPENADE, i.e. based on a Call-Graph whose nodes are indeed Flow-Graphs. This representation makes it easier to implement analyses in a way that is both context-sensitive and flow-sensitive. Our strategy also relies on the fixed-point implementation of the analyses, which uses a "wait-list".

At the same time, we continue the design of a adjoint-mode AD adapted to MPI communication. In our framework of AD by source transformation, we have pushed far in the direction of static data-flow analyses and static source transformation of individual MPI calls. We obtained results on classical cases of message-passing [38]. However, experience shows [11] that general usage of message-passing defies static analysis. A purely static analysis and transformation must resort too often to conservative choices, yielding a poor efficiency.

As a consequence, we are now going in the direction of a more dynamic, run-time treatment of adjoint MPI calls. This means designing a wrapper library "AMPI" on top of MPI, that takes care during execution of the adjoint code of the bookkeeping to send the adjoint messages in the reverse direction. This wrapper library should also be independent from the particular AD tool, as it will be used not only with TAPENADE but with the tools developed at Argonne and RWTH Aachen.

6.2. Finer control on AD transformation

Participants: Valérie Pascual, Laurent Hascoët.

We explore methods to provide the AD end-user with a better control on the AD transformation. We want to organize a progressive AD process in which the end-user can choose among a set of available AD code optimizations. In a first stage, the end-user may deactivate most of these optimizations, thus obtaining a differentiated code that is easier to understand and hopefully more robust. If problems do occur, this differentiated code is easier to debug with the debugging tools that we provide. In the next stages, the end-user may progressively turn the optimizations on, and at the same time check that the derivatives remain correct.

Another goal closely related is the comparison and evaluation of the existing corpus of AD code optimizations. TAPENADE is one of the AD tools that incorporate most of AD optimizations proposed in litterature. If a few missing optimizations are included, TAPENADE with its relatively large set of validation applications can be the common ground for a credible evaluation of the benefit brought by each optimization.

In this direction, we have extended TAPENADE to turn some classical optimizations that were automatically applied into optional optimizations. The emblematic example is activity analysis. This required some code cleanup. Also, we are extending TAPENADE to give the option of "association by address" instead of "association by name". This means bundling each variable with its derivative into a structured object, instead of creating new variables with new names to hold the derivatives. Which option is better is a difficult question, related to memory locality issues. This extension will allow us to make accurate measurements on our set of validation codes. This is also a step towards a better collaboration of TAPENADE with overloading-based AD tools, that natively use association by address.

6.3. Formal specification of AD

Participant: Laurent Hascoët.

There is very little formal specification of AD as a program transformation, and consequently no formal proof of its correctness. Correctness of tangent AD is problematic: if defined as equivalence of the tangent program semantics with the mathematical derivative of the semantics of the original code, correctness is mostly granted for simple staight-line programs, and in general not granted for programs with control. Therefore formal proofs of correctness appear unreachable at present. Fortunately, there is little concern about the practical relevance of tangent AD. The confidence of end-users regarding tangent AD is justified by everyday experience.

Adjoint-mode AD poses a different challenge. The adjoint AD transformation is by no means simple nor intuitive. Its specification is informal, so that end-users of AD cannot gain a strong confidence in the process. Moreover, the constant quest for efficiency of the adjoint code has introduced a number of improvements and tradeoffs that are defined informally. These improvements make the adjoint code intricate and sometimes interact to cause subtle bugs. On the other hand, the good news is that the difference between the adjoint code and the tangent code only lies in the order of the derivatives computations and not in their nature. A formal proof of semantic equivalence is thus conceivable.

The first step towards such a proof is a formal specification of both tangent-mode and adjoint-mode AD, including the specification of the program static data-flow analyses that the transformations require. We have provided this specification in terms of Data-Flow equations for the analyses, and in terms of Structural Operational Semantics (more precisely Natural Semantics) for the AD transformations themselves [19]. This specification will be the basis for future formal proofs of equivalence between tangent AD and adjoint AD.

6.4. Resolution of linearised systems

Participants: Hubert Alcin, Olivier Allain [Lemma], Marianna Braza [IMF-Toulouse], Alexandre Carabias, Alain Dervieux, Bruno Koobus [Université Montpellier 2], Carine Moussaed [Université Montpellier 2], Stephen Wornom [Lemma].

Increased sophistication of solution algorithms pose a challenge to Automatic Differentiation. Time-stepping iterations create numerous updates of the iterated solution vector. Other additional nonlinear iterative processes occur such as:

- the evaluation of an optimal step, which results at least from a homographic function of the unknown,
- the orthonormalisation of the updates (Gram-Schmidt method, Hessenberg method).

Adjoint-mode AD applied to these algorithms produces a “linearised iterative algorithm” which is transposed and therefore follows the original iterations in the reverse order, needing each of the iterated state solution vectors. One such extreme case is the simulation of unsteady phenomena with implicit numerical schemes: simulating high Reynolds turbulent flows by a Large Eddy Simulation (LES) and RANS-LES models requires hundreds of thousands time steps, each of them involving a modern iterative solution algorithm. This is the case targetted by the 4-year ANR project “ECINADS”, jointly with university of Montpellier 2, the Institut de Mécanique des fluides de Toulouse and Lemma company, started in 2009.

In ECINADS, we design more efficient solution algorithms and we examine the questions risen by their adjoint differentiation. Our goal is practical scalability of the direct simulation and of its adjoint on a large number of processors. ECINADS also addresses the scalable solution of new approximations.

In 2012, the novel three-level method studied by H. Alcin on a model problem has been extended to compressible viscous flows by B. Koobus and C. Moussaed from university of Montpellier.

Hubert Alcin, Bruno Koobus, Olivier Allain and Alain Dervieux published their work on a two-level Schwarz algorithm in IJNMFD [12]. H. Alcin has presented his work in the Parallel CFD conference of Altlanta [14]. H. Alcin wrote his thesis [11], defended in december, on the three main subjects of ECINADS: the two- and three-level Schwarz algorithms, Automatic Differentiation and mesh adaptation.

6.5. Automatic Differentiation of a CFD code

Participants: Hubert Alcin, Valérie Pascual, Laurent Hascoët, Alain Dervieux.

The ECINADS workplan includes the building of an adjoint state for a CFD kernel. We have chosen AIRONUM 5.1 , a real life kernel that combines two particular features:

- it uses intensively the Fortran95 dynamic memory allocation
- it uses MPI parallelization.

This work is reported in H. Alcin's PhD [11].

6.6. Perturbation Methods

Participants: Alain Dervieux, Laurent Hascoët.

In the context of the European project NODESIM-CFD (ended 2010), the contribution of Sciport involved mainly the derivation of perturbation methods and reduced order models for the management of uncertainties. These methods rely on Taylor series with second-order terms. The production of second derivative code is obtained through repeated application of Automatic Differentiation. Three strategies can be applied to obtain (elements of) the Hessian matrix, named Tangent-on-Tangent, Tangent-on-Adjoint, and Adjoint-on-Tangent. These new methods are promoted through short courses, e.g. by Alain Dervieux at an ERCOFTAC session (Chatou, 15-16 mai 2012). The application and extension of these methods are part of a FP7 proposal (Proposal UMRIDA, nov. 2012).

6.7. Control of approximation errors

Participants: Frédéric Alauzet [GAMMA team, Inria-Rocquencourt], Estelle Mbinky [GAMMA team, Inria-Rocquencourt], Olivier Allain [Lemma], Alexandre Carabias, Hubert Alcin, Alain Dervieux.

This is a joint research between Inria teams Gamma (Rocquencourt), Sciport, Castor and the Lemma company. Gamma brings mesh and approximation expertise, Sciport brings adjoint methods, and CFD applications are developed by CASTOR and Lemma.

The resolution of the optimum problem using adjoint-mode AD can be used in a slightly different context than optimal shape design, namely mesh adaptation. This will be possible if we can map the mesh adaptation problem into a differentiable optimal control problem. To this end, we express the mesh adaptation problem in a purely functional form: the mesh is reduced to a continuous property of the computational domain named the continuous metric. We minimize a continuous model of the error resulting from that metric. Thus the search of an adapted mesh is transformed into the search of an optimal metric.

In 2012, this activity is amplifying. A work on goal-oriented mesh adaptation for unsteady Euler flows submitted to the journal JCP has been accepted and published [13]. Its extension to the compressible Navier-Stokes model has been developed in 2D [22] and in 3D [11]. A further extension to Large Eddy Simulation has been defined and developed in the WOLF demonstrator. A communication at ECCOMAS (Vienna) has been presented and papers are being written for publication in journal.

The method is being extended to a third-order approximation, the Vertex-CENO. This approximation was defined collaboratively between university of Montpellier, IMM-Moscow and Sciport. A more accurate version is studied by Alexandre Carabias. A new mesh adaptation theory involving error estimates and criteria has been developed by Gamma and Sciport. The extension of the multiscale adaptation method is considered by Estelle Mbinky at Rocquencourt and has been presented at ECCOMAS (Vienna). The extension of the goal-oriented method is considered by Alexandre Carabias and first results were presented at ECCOMAS (Vienna). A cooperation with CEMEF and university of Nice is considered and a ERC common proposal, CMILE, has been built. Anisotropic mesh adaptation allows for better convergence towards continuous solutions, and in particular more accurate a posteriori error estimates and correctors. The synergy between correctors and mesh adaptation is the subject of a joint contribution (Gamma and Sciport) for the FP7 UMRIDA proposal (nov. 2012).

SELECT Project-Team

6. New Results

6.1. Model selection in Regression and Classification

Participants: Gilles Celeux, Mohammed El Anbari, Clément Levrard, Erwan Le Pennec, Lucie Montuelle, Pascal Massart, Caroline Meynet, Jean-Michel Poggi, Adrien Saumard.

Erwan Le Pennec is still working with Serge Cohen (IPANEMA Soleil) on hyperspectral image segmentation based on a spatialized Gaussian Mixture Model. Their scheme is supported by some theoretical investigation [6] and have been applied in pratice with an efficient minimization algorithm combining EM algorithm, dynamic programming and model selection implemented with MIXMOD. Lucie Montuelle is studying extensions of this model that comprise parametric logistic weights and regression mixtures.

In collaboration with Marie-Laure Martin-Magniette (URGV et UMR AgroParisTech/INRA MIA 518) and Cathy Maugis (INSA Toulouse) Gilles Celeux has extended their variable selection procedure for model-based clustering and supervised classification to deal with high dimensional data sets with a backward selection procedure which is more efficient that the previous forward selection procedure in this context. Moreover they have analysed the differences between the model-based approach and geometrical approach to select variable for clustering. Through numerical experiments, they showed the advantage of the model-based appraoch when many variables are highly correlated. These variable selection procedures are in particular used for genomics applications which is the result of a collaboration with researchers of of URGV (Evry Genopole).

Caroline Meynet provided an ℓ_1-oracle inequality satisfied by the Lasso estimator with the Kullback-Leibler loss in the framework of a finite mixture of Gaussian regressions model for high-dimensional heterogeneous data where the number of covariates may be much larger than the sample size. In particular, she has given a condition on the regularization parameter of the Lasso to obtain such an oracle inequality. This oracle inequality extends the ℓ_1-oracle inequality established by Massart and Meynet in the homogeneous Gaussian linear regression case. It is deduced from a finite mixture Gaussian regression model selection theorem for ℓ_1-penalized maximum likelihood conditional density estimation, which is inspired from Vapnik's method of structural risk minimization and from the theory on model selection for maximum likelihood estimators developed by Massart.

From an practical point of view, Caroline Meynet has introduced a procedure to select variables in model-based clustering in a high-dimensional context. In order to tackle with the problem of high-dimension, she has proposed to first use the Lasso in order to select different sets of variables and then estimate the density by a standard EM algorithm by reducing the inference to the linear space of the selected variables by the Lasso. Numerical experiments show that this method can outperform direct estimation by the Lasso.

In collaboration with Jean-Patrick Baudry (Paris 6) and Margarida Cardoso, Ana Ferreira and Maria-José Amorim (Lisbon University], Gilles Celeux has proposed an approach to select, in the model-based clustering context, a model and a number of clusters in order to get a partition which both provides a good fit with the data and is related to the external categorical variables. This approach makes use of the integrated joint likelihood of the data, the partition derived from the mixture model and the known partitions. It is worth noticing that the external categorical variables are only used to select a relevant mixture model. Each mixture model is fitted by the maximum likelihood methodology from the observed data. Numerical experiments illustrate the promising behaviour of the derived criterion [29].

Since September 2008, Pascal Massart is the cosupervisor with Frédéric Chazal (GEOMETRICA) of the thesis of Claire Caillerie (GEOMETRICA). The project intends to explore and to develop new researches at the crossing of information geometry, computational geometry and statistics.

Tim van Erven is studying Model Selection for the Long Term. When a model selection procedure forms an integrated part of a company's day-to-day activities, its performance should be measured not on a single day, but on average over a longer period, like for example a year. Taking this long-term perspective, it is possible to aggregate model predictions optimally even when the data probability distribution is so irregular that no statistical guarantees can be given for any individual day seperately. He studies the relation between model selection for individual days and for the long term, and how the geometry of the models affects both. This work has potential applications in model aggregation for the forecasting of electrical load consumption at EDF.

Adrien saumard has worked on the theoretical validation of the slope heuristics, a practical method of penalties calibration derived in a Gaussian setting by Birgé and Massart in 2006 and extended to bounded M-estimation by Arlot and Massart in 2010. He was able to prove the validity of this heuristics in bounded heteroscedastic regression with random design when the considered models where linear spans made of piecewise polynomials. A preliminary work on a fixed model was necessary and published in [9], while the validation of the slope heuristics itself - as well as the validation of a cross-validation approach - can be found in a preprint.

6.2. Statistical learning methodology and theory

Participants: Gilles Celeux, Christine Keribin, Erwan Le Pennec, Pascal Massart, Lucie Montuelle, Jean-Michel Poggi, Adrien Saumard, Solenne Thivin.

Unsupervised segmentation is an issue similar to unsupervised classification with an added spatial aspect. Functional data is acquired on points in a spatial domain and the goal is to segment the domain in homogeneous domain. The range of applications includes hyperspectral images in conservation sciences, fMRi data and all spatialized functional data. Erwan Le Pennec and Lucie Montuelle are focusing on the questions of the way to handle the spatial component from both the theoretical and the practical point of views. They study in particular the choice of the number of clusters. Furthermore, as functional data require heavy computation, they are required to propose numerically efficient algorithms. They have also extend the model to regression mixture.

Gilles Celeux, Christine Keribin and the Ph D. student Vincent Brault continue their work on the Latent Block Model (LBM). They compared several model selection criteria for binary tables [19]. However, the SEM-VEM Gibbs algorithm used to estimate LBM is subject to spurious solutions (empty clusters). To tackle this drawback, they have proposed to use Bayesian inference through Gibbs Sampling and studied the influence of the calibration of non informative prior distributions. They showed on numerical experiment the advantages of coupling Gibbs sampling with a Variational Bayes algorithm to get pointwise estimators [17]. Furthermore, they extended the previous studies from binary to categorical data [32].

Christine Keribin has proposed to compare, on genomics applications, the use of LBM with other methodologies (variable selection procedure of Maugis and Martin Magniette, component analysis). She supervised an internship (Master 1) on the use of principal component analysis for gene expression data (Inria funding). This has been done on data of the SONATA project (leaded by URGV - Evry Genopole), in collaboration with Marie-Laure Martin-Magniette.

Erwan Le Pennec is supervising Solenne Thivin in her CIFRE with Michel Prenat and Thales Optronique. The aim is target detection on complex background such as clouds or sea. Their approach is a local test approach based on the test decision theory. A key issue is to learn good discrimant features and their probabilistic properties. So far, they have worked on cloud images given by Thales. They focus on a Markovian modeling of the clouds.

Considering the case of maximum likelihood density estimation on histograms, Adrien saumard has investigated both theory and methodology. On the one hand, he has shown that AIC is twice the minimal penalty in the sense of Birgé and Massart, which by consequence implies the asymptotic optimality of the slope heuristics based on a linear shape. On the other hand, he investigated the methodology of the small to moderate sample size setting in this case. The robustness of the slope heurisitics compared to AIC is shown on simulated examples and a new overpenalization of Akaike's criterion is proposed, which outperforms the criterion AICc of

Hurvitch and Tsai and shows comparable results to the procedure proposed by Birgé and Rozenholc in 2006. The benefits of the derived procedure here is its theoretical background and interpretation. This work is still in process and some of the results can be found in a preprint.

6.3. Reliability and Computer Experiments

Participants: Yves Auffray, Gilles Celeux, Rémy Fouchereau, Shuai Fu.

In the computer experiments field, the goal is to approximate an expensive black box function from a limited number of evaluations. The choice of these evaluations i.e. the choice of a design of (computer) experiments is a major issue.

Following the previous work of the past three years, Shuai Fu has concluded her Ph.D thesis under the direction of Gilles Celeux [1]. This year, the work was focused on controlling four main error quantities, in order to validate the methodology in the industrial framework. More precisely, the DAC criterion (Data Agreement Criterion), which has been proposed for assessing the relevance of the design of experiments (DOE) and the prior choice with the observed data was applied to a complex hydrological model, coding and testing the relevant algorithms [30]. For the purpose of controlling the emulator error in an adaptive kriging algorithm, two Bayesian criteria have been proposed for searching and adding new points into the current DOE. The computation time remains important, which makes the method meaningful only in the case where we have a really time-consuming code.

In the framework of a CIFRE convention with Snecma-SAFRAN Rémy Fouchereau has started a thesis on the modeling of fatigue damage for Inco718 supervised by Gilles Celeux. Inco718 is a Zinc-based alloy. To determine its minimum lifetime, a lot of stress tests are made. The alloy lifetimes are reported as function of the stress. The aim is to propose a stochastic models for fatigue lifetime prediction based on a fracture mechanics-based approach. A mixture model with a lognormal component and a sum of two lognormals components is considered. Since the sum of two or more lognormal distribution is not closed form, inference on this model needs Monte Carlo integration within the EM algorithm. Thus, we have provided engineers with a probabilistic tool for reliability design of mechanical parts, but also with a diagnostic tool for material elaboration.

6.4. Statistical analysis of genomic data

Participant: Gilles Celeux.

In collaboration with Florence Jaffrezic and Andrea Rau (INRA, département de génétique animale) Gilles Celeux initiated modelling genomics networks from RNA-seq data. It was the subject of the internship of Mélina Gallpin who is starting a thesis on this subject. To day the performance of overdispersed Poisson models has been investigated. The results are somewhat poor especially for large numbers of genes.

6.5. Curves classification, denoising and forecasting

Participants: Émilie Devijver, Pascal Massart, Jean-Michel Poggi.

In collaboration with Farouk Mhamdi and Meriem Jaidane (ENIT, Tunis, Tunisia), Jean-Michel Poggi proposeda method for trend extraction from seasonal time series through the Empirical Mode Decomposition (EMD). Experimental comparison of trend extraction based on EMD, X11, X12 and Hodrick Prescott filter are conducted. First results show the eligibility of the blind EMD trend extraction method. Tunisian real peak load is also used to illustrate the extraction of the intrinsic trend.

In collaboration with Mina Aminghafari (Amirkabir University, Teheran), Jean-Michel Poggi made uses of wavelets in a statistical forecasting purpose for time series. Recent approaches involve wavelet decompositions in order to handle non stationary time series. They study and extended an approach proposed by Renaud et al., to estimate the prediction equation by direct regression of the process on the Haar non-decimated wavelet coefficients depending on its past values. The new variants are used first for stationary data and after for stationary data contaminated by a deterministic trend.

Jean-Michel Poggi was the supervisor (with A. Antoniadis) of the PhD Thesis of Jairo Cugliari-Duhalde which takes place in a CIFRE convention with EDF. It is strongly related to the use of wavelets together with curves clustering in order to perform accurate load comsumption forecasting. The thesis develops methodological and applied aspects linked to the electrical context as well as theoretical ones by introducing exogeneous variables in the context of nonparametric forecasting time series.

Jean-Michel Poggi, co-supervising with Anestis Antoniadis (Université Joseph Fourier Grenoble) the PhD thesis of Vincent Thouvenot, funded by a CIFRE with EDF. The industrial motivation of this work is the recent development of new technologies for measuring power consumption by EDF to acquire consumption data for different mesh network. The thesis will focus on the development of new statistical methods for predicting power consumption by exploiting the different levels of aggregation of network data collection. From the mathematical point of view, the work is to develop generalized additive models for this type of kind of aggregated data for the modeling of functional data, associating closely nonparametric estimation and variable selection using various penalization methods.

Jean-Michel Poggi and Pascal Massart are the co-advisors of the PhD thesis of Emilie Devijver, strongly motivated by the same kind of industrial forecasting problems in electricity, is dedicated to curves clustering for the prediction. A natural framework to explore this question is mixture of regression models for functional data. The theoretical subject of the thesis is to extend to functional data the recent work by Bühlmann et al. dealing with the simultaneous estimation of mixture regression models in the scalar case using Lasso type methods. Of course, it will be based on the technical tools of the work of Caroline Meynet (which completes his thesis Orsay under the direction of P. Massart), which deals with the clustering of functional data using Lasso methods choosing simultaneously number of clusters and selecting significant wavelet coefficients.

6.6. Neuroimaging, Statistical analysis of fMRI data

Participants: Gilles Celeux, Christine Keribin.

This research takes place as part of a collaboration with Neurospin on brain functional Magnetic Resonance Imaging (fMRI) data. (http://www.math.u-psud.fr/select/reunions/neurospin/Welcome.html). and concerns essentially regularisation in a supervised clustering methodology that includes spatial information in the prediction framework, and yields clustered weighted maps.

SEQUEL Project-Team

6. New Results

6.1. Decision-making Under Uncertainty

6.1.1. Reinforcement Learning

Transfer in Reinforcement Learning: a Framework and a Survey [56]

Transfer in reinforcement learning is a novel research area that focuses on the development of methods to transfer knowledge from a set of source tasks to a target task. Whenever the tasks are *similar*, the transferred knowledge can be used by a learning algorithm to solve the target task and significantly improve its performance (e.g., by reducing the number of samples needed to achieve a nearly optimal performance). In this chapter we provide a formalization of the general transfer problem, we identify the main settings which have been investigated so far, and we review the most important approaches to transfer in reinforcement learning.

Online Regret Bounds for Undiscounted Continuous Reinforcement Learning [44]

We derive sublinear regret bounds for undiscounted reinforcement learning in continuous state space. The proposed algorithm combines state aggregation with the use of upper confidence bounds for implementing optimism in the face of uncertainty. Beside the existence of an optimal policy which satisfies the Poisson equation, the only assumptions made are Holder continuity of rewards and transition probabilities.

Semi-Supervised Apprenticeship Learning [23]

In apprenticeship learning we aim to learn a good policy by observing the behavior of an expert or a set of experts. In particular, we consider the case where the expert acts so as to maximize an unknown reward function defined as a linear combination of a set of state features. In this paper, we consider the setting where we observe many sample trajectories (i.e., sequences of states) but only one or a few of them are labeled as experts' trajectories. We investigate the conditions under which the remaining unlabeled trajectories can help in learning a policy with a good performance. In particular, we define an extension to the max-margin inverse reinforcement learning proposed by Abbeel and Ng (2004) where, at each iteration, the max-margin optimization step is replaced by a semi-supervised optimization problem which favors classifiers separating clusters of trajectories. Finally, we report empirical results on two grid-world domains showing that the semi-supervised algorithm is able to output a better policy in fewer iterations than the related algorithm that does not take the unlabeled trajectories into account.

Fast Reinforcement Learning with Large Action Sets Using Error-Correcting Output Codes for MDP Factorization [31] [48]

The use of Reinforcement Learning in real-world scenarios is strongly limited by issues of scale. Most RL learning algorithms are unable to deal with problems composed of hundreds or sometimes even dozens of possible actions, and therefore cannot be applied to many real-world problems. We consider the RL problem in the supervised classification framework where the optimal policy is obtained through a multiclass classifier, the set of classes being the set of actions of the problem. We introduce error-correcting output codes (ECOCs) in this setting and propose two new methods for reducing complexity when using rollouts-based approaches. The first method consists in using an ECOC-based classifier as the multiclass classifier, reducing the learning complexity from O(A2) to O(Alog(A)) . We then propose a novel method that profits from the ECOC's coding dictionary to split the initial MDP into O(log(A)) separate two-action MDPs. This second method reduces learning complexity even further, from O(A2) to O(log(A)) , thus rendering problems with large action sets tractable. We finish by experimentally demonstrating the advantages of our approach on a set of benchmark problems, both in speed and performance.

Analysis of Classification-based Policy Iteration Algorithms [13]

We introduce a variant of the classification-based approach to policy iteration which uses a cost-sensitive loss function weighting each classification mistake by its actual regret, i.e., the difference between the action-value of the greedy action and of the action chosen by the classifier. For this algorithm, we provide a full finite-sample analysis. Our results state a performance bound in terms of the number of policy improvement steps, the number of rollouts used in each iteration, the capacity of the considered policy space (classifier), and a capacity measure which indicates how well the policy space can approximate policies that are greedy w.r.t. any of its members. The analysis reveals a tradeoff between the estimation and approximation errors in this classification-based policy iteration setting. Furthermore it confirms the intuition that classification-based policy iteration algorithms could be favorably compared to value-based approaches when the policies can be approximated more easily than their corresponding value functions. We also study the consistency of the algorithm when there exists a sequence of policy spaces with increasing capacity.

Minimax PAC-Bounds on the Sample Complexity of Reinforcement Learning with a Generative Model [5] [24]

We consider the problem of learning the optimal action-value function in discounted-reward Markov decision processes (MDPs). We prove new PAC bounds on the sample-complexity of two well-known model-based reinforcement learning (RL) algorithms in the presence of a generative model of the MDP: value iteration and policy iteration. The first result indicates that for an MDP with N state-action pairs and the discount factor $\gamma \in [0,1)$ only $O(N\log(N/\delta)/[(1-\gamma)^3\epsilon^2])$ state-transition samples are required to find an ϵ-optimal estimation of the action-value function with the probability (w.p.) $1-\delta$. Further, we prove that, for small values of ϵ, an order of $O(N\log(N/\delta)/[(1-\gamma)^3\epsilon^2])$ samples is required to find an ϵ-optimal policy w.p. $1-\delta$. We also prove a matching lower bound of $\Omega(N\log(N/\delta)/[(1-\gamma)^3\epsilon^2])$ on the sample complexity of estimating the optimal action-value function. To the best of our knowledge, this is the first minimax result on the sample complexity of RL: The upper bound matches the lower bound interms of N , ϵ, δ and $1/(1-\gamma)$ up to a constant factor. Also, both our lower bound and upper bound improve on the state-of-the-art in terms of their dependence on $1/(1-\gamma)$.

Optimistic planning in Markov decision processes [25]

The reinforcement learning community has recently intensified its interest in online planning methods, due to their relative independence on the state space size. However, tight near-optimality guarantees are not yet available for the general case of stochastic Markov decision processes and closed-loop, state-dependent planning policies. We therefore consider an algorithm related to $AO*$ that optimistically explores a tree representation of the space of closed-loop policies, and we analyze the near-optimality of the action it returns after n tree node expansions. While this optimistic planning requires a finite number of actions and possible next states for each transition, its asymptotic performance does not depend directly on these numbers, but only on the subset of nodes that significantly impact near-optimal policies. We characterize this set by introducing a novel measure of problem complexity, called the near-optimality exponent. Specializing the exponent and performance bound for some interesting classes of MDPs illustrates the algorithm works better when there are fewer near-optimal policies and less uniform transition probabilities.

Risk Bounds in Cost-sensitive Multiclass Classification: an Application to Reinforcement Learning [61]

We propose a computationally efficient classification-based policy iteration (CBPI) algorithm. The key idea of CBPI is to view the problem of computing the next policy in policy iteration as a classification problem. We propose a new cost-sensitive surrogate loss for each iteration of CBPI. This allows us to replace the non-convex optimization problem that needs to be solved at each iteration of the existing CBPI algorithms with a convex one. We show that the new loss is classification calibrated, and thus is a sound surrogate loss, and find a calibration function (i.e., a function that represents the convergence rate of the true loss in terms of the convergence rate of the surrogate-loss) for this loss. To the best of our knowledge, this is the first calibration result (with convergence rate) in the context of multi-class classification. As a result, we are able to extend the theoretical guarantees of the existing CBPI algorithms that deal with a non-convex optimization at each iteration to our convex and efficient algorithm, and thereby, obtain the first computationally efficient and theoretically sound CBPI algorithm.

Least-Squares Methods for Policy Iteration [55]

Approximate reinforcement learning deals with the essential problem of applying reinforcement learning in large and continuous state-action spaces, by us- ing function approximators to represent the solution. This chapter reviews least-squares methods for policy iteration, an important class of algorithms for approxi- mate reinforcement learning. We discuss three techniques for solving the core, pol- icy evaluation component of policy iteration, called: least-squares temporal difference, least-squares policy evaluation, and Bellman residual minimization. We introduce these techniques starting from their general mathematical principles and detailing them down to fully specified algorithms. We pay attention to online variants of policy iteration, and provide a numerical example highlighting the behavior of representative offline and online methods. For the policy evaluation component as well as for the overall resulting approximate policy iteration, we provide guarantees on the performance obtained asymptotically, as the number of processed samples and executed iterations grows to infinity. We also provide finite-sample results, which apply when a finite number of samples and iterations is considered. Finally, we outline several extensions and improvements to the techniques and methods reviewed

On Classification-based Approximate Policy Iteration [53]

Efficient methods for tackling large reinforcement learning problems usually exploit special structure, or regularities, of the problem at hand. For example, classification-based approximate policy iteration explicitly controls the complexity of the policy space, which leads to considerable improvement in convergence speed whenever the optimal policy is easy to represent. Conventional classification-based methods, however, do not benefit from regularities of the value function, because they typically use rollout-based estimates of the action-value function. This Monte Carlo-style approach for value estimation is data-inefficient and does not generalize the estimated value function over states. We introduce a general framework for classification-based approximate policy iteration (CAPI) which exploits regularities of both the policy and the value function. Our theoretical analysis extends existing work by allowing the policy evaluation step to be performed by any reinforcement learning algorithm (including temporal-difference style methods), by handling nonparametric representations of policies, and by providing tighter convergence bounds on the estimation error of policy learning. In our experiments, instantiations of CAPI outperformed powerful purely value-based approaches.

Conservative and Greedy Approaches to Classification-based Policy Iteration [37]

The existing classification-based policy iteration (CBPI) algorithms can be divided into two categories: *direct policy iteration* (DPI) methods that directly assign the output of the classifier (the approximate greedy policy w.r.t. the current policy) to the next policy, and *conservative policy iteration* (CPI) methods in which the new policy is a mixture distribution of the current policy and the output of the classifier. The conservative policy update gives CPI a desirable feature, namely the guarantee that the policies generated by this algorithm improve at each iteration. We provide a detailed algorithmic and theoretical comparison of these two classes of CBPI algorithms. Our results reveal that in order to achieve the same level of accuracy, CPI requires more iterations, and thus, more samples than the DPI algorithm. Furthermore, CPI may converge to suboptimal policies whose performance is not better than DPI's.

A Dantzig Selector Approach to Temporal Difference Learning [36]

LSTD is a popular algorithm for value function approximation. Whenever the number of features is larger than the number of samples, it must be paired with some form of regularization. In particular, l1-regularization methods tend to perform feature selection by promoting sparsity, and thus, are well- suited for high-dimensional problems. However, since LSTD is not a simple regression algorithm, but it solves a fixed-point problem, its integration with l1-regularization is not straightforward and might come with some drawbacks (e.g., the P-matrix assumption for LASSO-TD). In this paper, we introduce a novel algorithm obtained by integrating LSTD with the Dantzig Selector. We investigate the performance of the proposed algorithm and its relationship with the existing regularized approaches, and show how it addresses some of their drawbacks.

Finite-Sample Analysis of Least-Squares Policy Iteration [14]

In this paper, we report a performance bound for the widely used least-squares policy iteration (LSPI) algorithm. We first consider the problem of policy evaluation in reinforcement learning, that is, learning the value function of a fixed policy, using the least-squares temporal-difference (LSTD) learning method, and report finite-sample analysis for this algorithm. To do so, we first derive a bound on the performance of the LSTD solution evaluated at the states generated by the Markov chain and used by the algorithm to learn an estimate of the value function. This result is general in the sense that no assumption is made on the existence of a stationary distribution for the Markov chain. We then derive generalization bounds in the case when the Markov chain possesses a stationary distribution and is β-mixing. Finally, we analyze how the error at each policy evaluation step is propagated through the iterations of a policy iteration method, and derive a performance bound for the LSPI algorithm.

Approximate Modified Policy Iteration [47]

Modified policy iteration (MPI) is a dynamic programming (DP) algorithm that contains the two celebrated policy and value iteration methods. Despite its generality, MPI has not been thoroughly studied, especially its approximation form which is used when the state and/or action spaces are large or infinite. In this paper, we propose three implementations of approximate MPI (AMPI) that are extensions of well-known approximate DP algorithms: fitted-value iteration, fitted-Q iteration, and classification-based policy iteration. We provide error propagation analyses that unify those for approximate policy and value iteration. On the last classification-based implementation, we develop a finite-sample analysis that shows that MPI's main parameter allows to control the balance between the estimation error of the classifier and the overall value function approximation.

Bayesian Reinforcement Learning [57]

This chapter surveys recent lines of work that use Bayesian techniques for reinforcement learning. In Bayesian learning, uncertainty is expressed by a prior distribution over unknown parameters and learning is achieved by computing a posterior distribution based on the data observed. Hence, Bayesian reinforcement learning distinguishes itself from other forms of reinforcement learning by explicitly maintaining a distribution over various quantities such as the parameters of the model, the value function, the policy or its gradient. This yields several benefits: a) domain knowledge can be naturally encoded in the prior distribution to speed up learning; b) the exploration/exploitation tradeoff can be naturally optimized; and c) notions of risk can be naturally taken into account to obtain robust policies.

6.1.2. Multi-arm Bandit Theory

Learning with stochastic inputs and adversarial outputs [15]

Most of the research in online learning is focused either on the problem of adversarial classification (i.e., both inputs and labels are arbitrarily chosen by an adversary) or on the traditional supervised learning problem in which samples are independent and identically distributed according to a stationary probability distribution. Nonetheless, in a number of domains the relationship between inputs and outputs may be adversarial, whereas input instances are i.i.d. from a stationary distribution (e.g., user preferences). This scenario can be formalized as a learning problem with stochastic inputs and adversarial outputs. In this paper, we introduce this novel stochastic-adversarial learning setting and we analyze its learnability. In particular, we show that in a binary classification problem over an horizon of n rounds, given a hypothesis space H with finite VC-dimension, it is possible to design an algorithm that incrementally builds a suitable finite set of hypotheses from H used as input for an exponentially weighted forecaster and achieves a cumulative regret of order $O(\sqrt{nVC(H)\log n})$ with overwhelming probability. This result shows that whenever inputs are i.i.d. it is possible to solve any binary classification problem using a finite VC-dimension hypothesis space with a sub-linear regret independently from the way labels are generated (either stochastic or adversarial). We also discuss extensions to multi-class classification, regression, learning from experts and bandit settings with stochastic side information, and application to games.

A Truthful Learning Mechanism for Multi-Slot Sponsored Search Auctions with Externalities [35]

Sponsored search auctions constitute one of the most successful applications of *microeconomic mechanisms*. In mechanism design, auctions are usually designed to incentivize advertisers to bid their truthful valuations and, at the same time, to assure both the advertisers and the auctioneer a non–negative utility. Nonetheless, in sponsored search auctions, the click–through–rates (CTRs) of the advertisers are often unknown to the auctioneer and thus standard incentive compatible mechanisms cannot be directly applied and must be paired with an effective learning algorithm for the estimation of the CTRs. This introduces the critical problem of designing a learning mechanism able to estimate the CTRs as the same time as implementing a truthful mechanism with a revenue loss as small as possible compared to an optimal mechanism designed with the true CTRs. Previous works showed that in single–slot auctions the problem can be solved using a suitable exploration–exploitation mechanism able to achieve a per–step regret of order $O(T^{-1/3})$ (where T is the number of times the auction is repeated). In this paper we extend these results to the general case of contextual multi–slot auctions with position– and ad–dependent externalities. In particular, we prove novel upper–bounds on the revenue loss w.r.t. to a VCG auction and we report numerical simulations investigating their accuracy in predicting the dependency of the regret on the number of rounds T, the number of slots K, and the number of advertisements n.

Regret Bounds for Restless Markov Bandits [43]

We consider the restless Markov bandit problem, in which the state of each arm evolves according to a Markov process independently of the learner's actions. We suggest an algorithm that after T steps achieves $\widetilde{O}(\sqrt{T})$ regret with respect to the best policy that knows the distributions of all arms. No assumptions on the Markov chains are made except that they are irreducible. In addition, we show that index-based policies are necessarily suboptimal for the considered problem.

Online allocation and homogeneous partitioning for piecewise constant mean approximation [42]

In the setting of active learning for the multi-armed bandit, where the goal of a learner is to estimate with equal precision the mean of a finite number of arms, recent results show that it is possible to derive strategies based on finite-time confidence bounds that are competitive with the best possible strategy. We here consider an extension of this problem to the case when the arms are the cells of a finite partition P of a continuous sampling space X in Rd. Our goal is now to build a piecewise constant approximation of a noisy function (where each piece is one region of P and P is fixed beforehand) in order to maintain the local quadratic error of approximation on each cell equally low. Although this extension is not trivial, we show that a simple algorithm based on upper confidence bounds can be proved to be adaptive to the function itself in a near-optimal way, when |P| is chosen to be of minimax-optimal order on the class of alpha-Holder functions.

The Optimistic Principle applied to Games, Optimization and Planning: Towards Foundations of Monte-Carlo Tree Search [17]

This work covers several aspects of the optimism in the face of uncertainty principle applied to large scale optimization problems under finite numerical budget. The initial motivation for the research reported here originated from the empirical success of the so-called Monte-Carlo Tree Search method popularized in computer-go and further extended to many other games as well as optimization and planning problems. Our objective is to contribute to the development of theoretical foundations of the field by characterizing the complexity of the underlying optimization problems and designing efficient algorithms with performance guarantees. The main idea presented here is that it is possible to decompose a complex decision making problem (such as an optimization problem in a large search space) into a sequence of elementary decisions, where each decision of the sequence is solved using a (stochastic) multi-armed bandit (simple mathematical model for decision making in stochastic environments). This so-called hierarchical bandit approach (where the reward observed by a bandit in the hierarchy is itself the return of another bandit at a deeper level) possesses the nice feature of starting the exploration by a quasi-uniform sampling of the space and then focusing progressively on the most promising area, at different scales, according to the evaluations observed so far, and eventually performing a local search around the global optima of the function. The performance of the method is assessed in terms of the optimality of the returned solution as a function of the number of function evaluations. Our main contribution to the field of function optimization is a class of hierarchical optimistic algorithms designed for general search spaces (such as metric spaces, trees, graphs, Euclidean spaces, ...) with different algorithmic instantiations depending on whether the evaluations are noisy or noiseless and whether some measure of the ”smoothness” of the function is known or unknown. The performance of the algorithms depend on the local behavior of the function around its global optima expressed in terms of the quantity of near-optimal states measured with some metric. If this local smoothness of the function is known then one can design very efficient optimization algorithms (with convergence rate independent of the space dimension), and when it is not known, we can build adaptive techniques that can, in some cases, perform almost as well as when it is known.

Kullback-Leibler Upper Confidence Bounds for Optimal Sequential Allocation [6]

We consider optimal sequential allocation in the context of the so-called stochastic multi-armed bandit model. We describe a generic index policy, in the sense of Gittins (1979), based on upper confidence bounds of the arm payoffs computed using the Kullback-Leibler divergence. We consider two classes of distributions for which instances of this general idea are analyzed: The kl-UCB algorithm is designed for one-parameter exponential families and the empirical KL-UCB algorithm for bounded and finitely supported distributions. Our main contribution is a unified finite-time analysis of the regret of these algorithms that asymptotically matches the lower bounds of Lai and Robbins (1985) and Burnetas and Katehakis (1996), respectively. We also investigate the behavior of these algorithms when used with general bounded rewards, showing in particular that they provide significant improvements over the state-of-the-art.

Minimax strategy for Stratified Sampling for Monte Carlo [8]

We consider the problem of stratified sampling for Monte-Carlo integration. We model this problem in a multi-armed bandit setting, where the arms represent the strata, and the goal is to estimate a weighted average of the mean values of the arms. We propose a strategy that samples the arms according to an upper bound on their standard deviations and compare its estimation quality to an ideal allocation that would know the standard deviations of the strata. We provide two pseudo-regret analyses: a distribution-dependent bound of order $O(n^{-3/2})$ that depends on a measure of the disparity of the strata, and a distribution-free bound $O(n^{-4/3})$ that does not. We also provide the first problem independent (minimax) lower bound for this problem and demonstrate that MC-UCB matches this lower bound both in terms of number of samples n and in terms of number of strata K. Finally, we link the pseudo-regret with the difference between the mean squared error on the estimated weighted average of the mean values of the arms, and the optimal oracle strategy: this provides us also with a problem dependent and a problem independent rate for this measure of performance and, as a corollary, asymptotic optimality.

Upper-Confidence-Bound Algorithms for Active Learning in Multi-Armed Bandits [7]

In this paper, we study the problem of estimating uniformly well the mean values of several distributions given a finite budget of samples. If the variance of the distributions were known, one could design an optimal sampling strategy by collecting a number of independent samples per distribution that is proportional to their variance. However, in the more realistic case where the distributions are not known in advance, one needs to design adaptive sampling strategies in order to select which distribution to sample from according to the previously observed samples. We describe two strategies based on pulling the distributions a number of times that is proportional to a high-probability upper-confidence-bound on their variance (built from previous observed samples) and report a finite-sample performance analysis on the excess estimation error compared to the optimal allocation. We show that the performance of these allocation strategies depends not only on the variances but also on the full shape of the distributions.

Bandit Algorithms boost motor-task selection for Brain Computer Interfaces [32] [10]

Brain-computer interfaces (BCI) allow users to “communicate” with a computer without using their muscles. BCI based on sensori-motor rhythms use imaginary motor tasks, such as moving the right or left hand, to send control signals. The performances of a BCI can vary greatly across users but also depend on the tasks used, making the problem of appropriate task selection an important issue. This study presents a new procedure to automatically select as fast as possible a discriminant motor task for a brain-controlled button. We develop for this purpose an adaptive algorithm, *UCB-classif*, based on the stochastic bandit theory. This shortens the training stage, thereby allowing the exploration of a greater variety of tasks. By not wasting time on inefficient tasks, and focusing on the most promising ones, this algorithm results in a faster task selection and a more efficient use of the BCI training session. Comparing the proposed method to the standard practice in task selection, for a fixed time budget, *UCB-classif* leads to an improved classification rate, and for a fixed classification rate, to a reduction of the time spent in training by 50%.

Adaptive Stratified Sampling for Monte-Carlo integration of Differentiable functions [26]

We consider the problem of adaptive stratified sampling for Monte Carlo integration of a differentiable function given a finite number of evaluations to the function. We construct a sampling scheme that samples more often in regions where the function oscillates more, while allocating the samples such that they are well spread on the domain (this notion shares similitude with low discrepancy). We prove that the estimate returned by the algorithm is almost similarly accurate as the estimate that an optimal oracle strategy (that would know the variations of the function *everywhere*) would return, and provide a finite-sample analysis.

Risk-Aversion in Multi-Armed Bandits [46]

In stochastic multi-armed bandits the objective is to solve the exploration-exploitation dilemma and ultimately maximize the expected reward. Nonetheless, in many practical problems, maximizing the expected reward is not the most desirable objective. In this paper, we introduce a novel setting based on the principle of risk-aversion where the objective is to compete against the arm with the best risk-return trade-off. This setting proves to be intrinsically more difficult than the standard multi-arm bandit setting due in part to an exploration risk which introduces a regret associated to the variability of an algorithm. Using variance as a measure of risk, we introduce two new algorithms, we investigate their theoretical guarantees, and we report preliminary empirical results.

Bandit Theory meets Compressed Sesing for high dimensional Stochastic Linear Bandit [27]

We consider a linear stochastic bandit problem where the dimension K of the unknown parameter θ is larger than the sampling budget n. In such cases, it is in general impossible to derive sub-linear regret bounds since usual linear bandit algorithms have a regret in $O(K\sqrt{n})$. In this paper we assume that θ is S-sparse, i.e. has at most S non-zero components, and that the space of arms is the unit ball for the L_2 norm. We combine ideas from Compressed Sensing and Bandit Theory and derive an algorithm with a regret bound in $O(S\sqrt{n})$. We detail an application to the problem of optimizing a function that depends on many variables but among which only a small number of them (initially unknown) are relevant.

Thompson Sampling: an Asymptotically Optimal Finite Time Analysis [38]

The question of the optimality of Thompson Sampling for solving the stochastic multi-armed bandit problem had been open since 1933. In this paper we answer it positively for the case of Bernoulli rewards by providing the first finite-time analysis that matches the asymptotic rate given in the Lai and Robbins lower bound for the cumulative regret. The proof is accompanied by a numerical comparison with other optimal policies, experiments that have been lacking in the literature until now for the Bernoulli case.

Regret bounds for Restless Markov Bandits [43]

We consider the restless Markov bandit problem, in which the state of each arm evolves according to a Markov process independently of the learner's actions. We suggest an algorithm that after T steps achieves $O(\sqrt{T})$ regret with respect to the best policy that knows the distributions of all arms. No assumptions on the Markov chains are made except that they are irreducible. In addition, we show that index-based policies are necessarily suboptimal for the considered problem.

Minimax number of strata for online Stratified Sampling given Noisy Samples [28]

We consider the problem of online stratified sampling for Monte Carlo integration of a function given a finite budget of n noisy evaluations to the function. More precisely we focus on the problem of choosing the number of strata K as a function of the budget n. We provide asymptotic and finite-time results on how an oracle that has access to the function would choose the number of strata optimally. In addition we prove a lower bound on the learning rate for the problem of stratified Monte-Carlo. As a result, we are able to state, by improving the bound on its performance, that algorithm MC-UCB, is minimax optimal both in terms of the number of samples n and the number of strata K, up to a $\log(nK)$ factor. This enables to deduce a minimax optimal bound on the difference between the performance of the estimate output by MC-UCB, and the performance of the estimate output by the best oracle static strategy, on the class of Holder continuous functions, and up to a factor $\log(n)$.

Best Arm Identification: A Unified Approach to Fixed Budget and Fixed Confidence [33]

We study the problem of identifying the best arm(s) in the stochastic multi-armed bandit setting. This problem has been studied in the literature from two different perspectives: fixed budget and fixed confidence. We propose a unifying approach that leads to a meta-algorithm called unified gap-based exploration (UGapE), with a common structure and similar theoretical analysis for these two settings. We prove a performance bound for the two versions of the algorithm showing that the two problems are characterized by the same notion of complexity. We also show how the UGapE algorithm as well as its theoretical analysis can be extended to take into account the variance of the arms and to multiple bandits. Finally, we evaluate the performance of UGapE and compare it with a number of existing fixed budget and fixed confidence algorithms.

6.2. Statistical Analysis of Time Series

6.2.1. Prediction of Sequences of Structured and Unstructured Data

Reducing statistical time-series problems to binary classification [45]

We show how binary classification methods developed to work on i.i.d. data can be used for solving statistical problems that are seemingly unrelated to classification and concern highly-dependent time series. Specifically, the problems of time-series clustering, homogeneity testing and the three-sample problem are addressed. The algorithms that we construct for solving these problems are based on a new metric between time-series distributions, which can be evaluated using binary classification methods. Universal consistency of the proposed algorithms is proven under most general assumptions. The theoretical results are illustrated with experiments on synthetic and real-world data.

6.2.2. Hypothesis Testing

Testing composite hypotheses about discrete ergodic processes [21]

Given a discrete-valued sample $X_1, \cdots, X_n$ we wish to decide whether it was generated by a distribution belonging to a family H_0, or it was generated by a distribution belonging to a family H_1. In this work we assume that all distributions are stationary ergodic, and do not make any further assumptions (in particular, no independence or mixing rate assumptions). We find some necessary and some sufficient conditions, formulated in terms of the topological properties of H_0 and H_1, for the existence of a consistent test. For the case when H_1 is the complement of H_0 (to the set of all stationary ergodic processes) these necessary and sufficient conditions coincide, thereby providing a complete characterization of families of processes membership to which can be consistently tested, against their complement, based on sampling. This criterion includes as special cases several known and some new results on testing for membership to various parametric families, as well as testing identity, independence, and other hypotheses.

Uniform hypothesis testing for finite-valued stationary processes [22]

Given a discrete-valued sample $X_1, \cdots, X_n$ we wish to decide whether it was generated by a distribution belonging to a family H_0, or it was generated by a distribution belonging to a family H_1. In this work we assume that all distributions are stationary ergodic, and do not make any further assumptions (e.g. no independence or mixing rate assumptions). We would like to have a test whose probability of error (both Type I and Type II) is uniformly bounded. More precisely, we require that for each ϵ there exist a sample size n such that probability of error is upper-bounded by ϵ for samples longer than n. We find some necessary and some sufficient conditions on H_0 and H_1 under which a consistent test (with this notion of consistency) exists. These conditions are topological, with respect to the topology of distributional distance.

6.2.3. Change Point Analysis

Locating Changes in Highly Dependent Data with Unknown Number of Change Points [39]

The problem of multiple change point estimation is considered for sequences with unknown number of change points. A consistency framework is suggested that is suitable for highly dependent time-series, and an asymptotically consistent algorithm is proposed. In order for the consistency to be established the only assumption required is that the data is generated by stationary ergodic time-series distributions. No modeling, independence or parametric assumptions are made; the data are allowed to be dependent and the dependence can be of arbitrary form. The theoretical results are complemented with experimental evaluations.

6.2.4. Clustering Time Series, Online and Offline

Online Clustering of Processes [40]

The problem of online clustering is considered in the case where each data point is a sequence generated by a stationary ergodic process. Data arrive in an online fashion so that the sample received at every time-step is either a continuation of some previously received sequence or a new sequence. The dependence between the sequences can be arbitrary. No parametric or independence assumptions are made; the only assumption is that the marginal distribution of each sequence is stationary and ergodic. A novel, computationally efficient algorithm is proposed and is shown to be asymptotically consistent (under a natural notion of consistency). The performance of the proposed algorithm is evaluated on simulated data, as well as on real datasets (motion classification).

Incremental Spectral Clustering with the Normalised Laplacian [52]

Partitioning a graph into groups of vertices such that those within each group are more densely connected than vertices assigned to different groups, known as graph clustering, is often used to gain insight into the organization of large scale networks and for visualization purposes. Whereas a large number of dedicated techniques have been recently proposed for static graphs, the design of on-line graph clustering methods tailored for evolving networks is a challenging problem, and much less documented in the literature. Motivated by the broad variety of applications concerned, ranging from the study of biological networks to graphs of scientific references through to the exploration of communications networks such as the World Wide Web, it is the main purpose of this paper to introduce a novel, computationally efficient, approach to graph clustering in the evolutionary context. Namely, the method promoted in this article is an incremental eigenvalue solution for the spectral clustering method described by Ng. et al. (2001). Beyond a precise description of its practical implementation and an evaluation of its complexity, its performance is illustrated through numerical experiments, based on datasets modelling the evolution of a HIV epidemic and the purchase history graph of an e-commerce website.

6.2.5. Online Semi-Supervised Learning

Learning from a Single Labeled Face and a Stream of Unlabeled Data [41]

Face recognition from a single image per person is a challenging problem because the training sample is extremely small. We consider a variation of this problem. In our problem, we recognize only one person, and there are no labeled data for any other person. This setting naturally arises in authentication on personal computers and mobile devices, and poses additional challenges because it lacks negative examples. We formalize our problem as one-class classification, and propose and analyze an algorithm that learns a non-parametric model of the face from a single labeled image and a stream of unlabeled data. In many domains, for instance when a person interacts with a computer with a camera, unlabeled data are abundant and easy to utilize. This is the first paper that investigates how these data can help in learning better models in the single-image-per-person setting. Our method is evaluated on a dataset of 43 people and we show that these people can be recognized 90% of time at nearly zero false positives. This recall is 25+% higher than the recall of our best performing baseline. Finally, we conduct a comprehensive sensitivity analysis of our algorithm and provide a guideline for setting its parameters in practice.

6.3. Statistical Learning and Bayesian Analysis

6.3.1. Non-parametric Methods for Function Approximation

Linear Regression with Random Projections [16]

We investigate a method for regression that makes use of a randomly generated subspace G_P (of finite dimension P) of a given large (possibly infinite) dimensional function space F, for example, $L_2([0,1]^d)$. G_P is defined as the span of P random features that are linear combinations of a basis functions of F weighted by random Gaussian i.i.d. coefficients. We show practical motivation for the use of this approach, detail the link that this random projections method share with RKHS and Gaussian objects theory and prove, both in deterministic and random design, approximation error bounds when searching for the best regression function in G_P rather than in F, and derive excess risk bounds for a specific regression algorithm (least squares regression in G_P). This paper stresses the motivation to study such methods, thus the analysis developed is kept simple for explanations purpose and leaves room for future developments.

6.3.2. Nonparametric Bayesian Estimation

DPM pour l'inférence dans les modèles dynamiques non linéaires avec des bruits de mesure alpha-stable [50]

Stable random variables are often use to model impulsive noise; Recently it has be shown that communication at very high frequency suffer from such a noise. Stable noise cannot however be considered as usual noise in estimation processes because the variance does not usually exists nor an analytic expression for the probability density function. In this work we show how to manage such a problem using a bayesian nonparametric approach. We develop a Sequential Monte Carlo based algorithm to realize the estimation in a non linear dynamical system. The measurement noise is a non-stationnary stable process and it is modeled using a Dirichlet Process Mixture.

6.3.3. Random Finite Sets for Multisensor Multitarget Tracking

Multi-sensor PHD filtering with application to sensor management [2]

The aim of multi-object filtering is to address the multiple target detection and/or tracking problem. This thesis focuses on the Probability Hypothesis Density (PHD) filter, a well-known tractable approximation of the Random Finite Set (RFS) filter when the observation process is realized by a single sensor. The first part proposes the rigorous construction of the exact multi-sensor PHD filter and its simplified expression, without approximation, through a joint partitioning of the target state space and the sensors. With this new method, the exact multi-sensor PHD can be propagated in simple surveillance scenarii. The second part deals with the sensor management problem in the PHD framework. At each iteration, the Balanced Explorer and Tracker (BET) builds a prediction of the posterior multi-sensor PHD thanks to the Predicted Ideal Measurement Set (PIMS) and produces a multi-sensor control according to a few simple operational principles adapted to surveillance activities

6.4. Applications

6.4.1. Signal Processing

Dirichlet Process Mixtures for Density Estimation in Dynamic Nonlinear Modeling: Application to GPS Positioning in Urban Canyons [19]

In global positioning systems (GPS), classical localization algorithms assume, when the signal is received from the satellite in line-of-sight (LOS) environment, that the pseudorange error distribution is Gaussian. Such assumption is in some way very restrictive since a random error in the pseudorange measure with an unknown distribution form is always induced in constrained environments especially in urban canyons due to multipath/masking effects. In order to ensure high accuracy positioning, a good estimation of the observation error in these cases is required. To address this, an attractive flexible Bayesian nonparametric noise model based on Dirichlet process mixtures (DPM) is introduced. Since the considered positioning problem involves elements of non-Gaussianity and nonlinearity and besides, it should be processed on-line, the suitability of the proposed modeling scheme in a joint state/parameter estimation problem is handled by an efficient Rao-Blackwellized particle filter (RBPF). Our approach is illustrated on a data analysis task dealing with joint estimation of vehicles positions and pseudorange errors in a global navigation satellite system (GNSS)-based localization context where the GPS information may be inaccurate because of hard reception conditions.

Dislocation detection in field environments: A belief functions contribution [20]

Dislocation is defined as the change between discrete sequential locations of critical items in field environments such as large construction projects. Dislocations on large sites of materials and critical items for which discrete time position estimates are available represent critical state changes. The ability to detect dislocations automatically for tens of thousands of items can ultimately improve project performance significantly. Detecting these dislocations in a noisy information environment where low cost radio frequency identification tags are attached to each piece of material, and the material is moved sometimes only a few meters, is the main focus of this study. We propose in this paper a method developed in the frame of belief functions to detect dislocations. The belief function framework is well-suited for such a problem where both uncertainty and imprecision are inherent to the problem. We also show how to deal with the calculations. This method has been implemented in a controlled experimental setting. The results of these experiments show the ability of the proposed method to detect materials dislocation over the site reliably. Broader application of this approach to both animate and inanimate objects is possible.

Towards dictionary learning from images with non Gaussian noise [29]

We address the problem of image dictionary learning from noisy images with non Gaussian noise. This problem is difficult. As a first step, we consider the extreme sparse code given by vector quantization, i.e. each pixel is finally associated to 1 single atom. For Gaussian noise, the natural solution is K-means clustering using the sum of the squares of differences between gray levels as the dissimilarity measure between patches. For non Gaussian noises (Poisson, Gamma,...), a new measure of dissimilarity between noisy patches is necessary. We study the use of the generalized likelihood ratios (GLR) recently introduced by Deledalle et al. 2012 to compare non Gaussian noisy patches. We propose a K-medoids algorithm generalizing the usual Linde-Buzo-Gray K-means using the GLR based dissimilarity measure. We obtain a vector quantization which provides a dictionary that can be very large and redundant. We illustrate our approach by dictionaries learnt from images featuring non Gaussian noise, and present preliminary denoising results.

6.4.2. Medical Applications

Outlier detection for patient monitoring and alerting. [12]

We develop and evaluate a data-driven approach for detecting unusual (anomalous) patient-management decisions using past patient cases stored in electronic health records (EHRs). Our hypothesis is that a patient-management decision that is unusual with respect to past patient care may be due to an error and that it is worthwhile to generate an alert if such a decision is encountered. We evaluate this hypothesis using data obtained from EHRs of 4486 post-cardiac surgical patients and a subset of 222 alerts generated from the data. We base the evaluation on the opinions of a panel of experts. The results of the study support our hypothesis that the outlier-based alerting can lead to promising true alert rates. We observed true alert rates that ranged from 25% to 66% for a variety of patient-management actions, with 66% corresponding to the strongest outliers.

6.4.3. Web Mining

Managing advertising campaigns – an approximate planning approach [11]

We consider the problem of displaying commercial advertisements on web pages, in the "cost per click" model. The advertisement server has to learn the appeal of each type of visitor for the different advertisements in order to maximize the profit. Advertisements have constraints such as a certain number of clicks to draw, as well as a lifetime. This problem is thus inherently dynamic, and intimately combines combinatorial and statistical issues. To set the stage, it is also noteworthy that we deal with very rare events of interest, since the base probability of one click is in the order of 10^4. Different approaches may be thought of, ranging from computationally demanding ones (use of Markov decision processes, or stochastic programming) to very fast ones.We introduce NOSEED, an adaptive policy learning algorithm based on a combination of linear programming and multi-arm bandits. We also propose a way to evaluate the extent to which we have to handle the constraints (which is directly related to the computation cost). We investigate the performance of our system through simulations on a realistic model designed with an important commercial web actor.

ICML Exploration & Exploitation challenge: Keep it simple! [18]

Recommendation has become a key feature in the economy of a lot of companies (online shopping, search engines...). There is a lot of work going on regarding recommender systems and there is still a lot to do to improve them. Indeed nowadays in many companies most of the job is done by hand. Moreover even when a supposedly smart recommender system is designed, it is hard to evaluate it without using real audience which obviously involves economic issues. The ICML Exploration & Exploitation challenge is an attempt to make people propose efficient recommendation techniques and particularly focuses on limited computational resources. The challenge also proposes a framework to address the problem of evaluating a recommendation algorithm with real data. We took part in this challenge and achieved the best performances; this paper aims at reporting on this achievement; we also discuss the evaluation process and propose a better one for future challenges of the same kind.

6.4.4. Games

CLOP: Confident Local Optimization for Noisy Black-Box Parameter Tuning [30]

Artificial intelligence in games often leads to the problem of parameter tuning. Some heuristics may have coefficients, and they should be tuned to maximize the win rate of the program. A possible approach is to build local quadratic models of the win rate as a function of program parameters. Many local regression algorithms have already been proposed for this task, but they are usually not robust enough to deal automatically and efficiently with very noisy outputs and non-negative Hessians. The CLOP principle, which stands for Confident Local OPtimization, is a new approach to local regression that overcomes all these problems in a simple and efficient way. CLOP discards samples whose estimated value is confidently inferior to the mean of all samples. Experiments demonstrate that, when the function to be optimized is smooth, this method outperforms all other tested algorithms.

6.5. Other Results

Sequential approaches for learning datum-wise sparse representations [9]

In supervised classification, data representation is usually considered at the dataset level: one looks for the "best" representation of data assuming it to be the same for all the data in the data space. We propose a different approach where the representations used for classification are tailored to each datum in the data space. One immediate goal is to obtain sparse datum-wise representations: our approach learns to build a representation specific to each datum that contains only a small subset of the features, thus allowing classification to be fast and efficient. This representation is obtained by way of a sequential decision process that sequentially chooses which features to acquire before classifying a particular point; this process is learned through algorithms based on Reinforcement Learning. The proposed method performs well on an ensemble of medium-sized sparse classification problems. It offers an alternative to global sparsity approaches, and is a natural framework for sequential classification problems. The method extends easily to a whole family of sparsity-related problems which would otherwise require developing specific solutions. This is the case in particular for cost-sensitive and limited-budget classification, where feature acquisition is costly and is often performed sequentially. Finally, our approach can handle non-differentiable loss functions or combinatorial optimization encountered in more complex feature selection problems.

Multiple Operator-valued Kernel Learning [60]

Positive definite operator-valued kernels generalize the well-known notion of reproducing kernels, and are naturally adapted to multi-output learning situations. This paper addresses the problem of learning a finite linear combination of infinite-dimensional operator-valued kernels which are suitable for extending functional data analysis methods to nonlinear contexts. We study this problem in the case of kernel ridge regression for functional responses with an lr-norm constraint on the combination coefficients. The resulting optimization problem is more involved than those of multiple scalar-valued kernel learning since operator-valued kernels pose more technical and theoretical issues. We propose a multiple operator-valued kernel learning algorithm based on solving a system of linear operator equations by using a block coordinatedescent procedure. We experimentally validate our approach on a functional regression task in the context of finger movement prediction in brain-computer interfaces.

SIERRA Project-Team

6. New Results

6.1. A Stochastic Gradient Method with an Exponential Convergence Rate for Strongly-Convex Optimization with Finite Training Sets

Participants: Francis Bach, Mark Schmidt, Nicolas Le Roux [correspondant].

In [21], we propose a new stochastic gradient method for optimizing the sum of a finite set of smooth functions, where the sum is strongly convex. While standard stochastic gradient methods converge at sublinear rates for this problem, the proposed method incorporates a memory of previous gradient values in order to achieve a linear convergence rate. In a machine learning context, numerical experiments indicate that the new algorithm can dramatically outperform standard algorithms, both in terms of optimizing the training objective and reducing the testing objective quickly.

6.2. Convex Relaxation for Combinatorial Penalties

Participants: Francis Bach, Guillaume Obozinski [correspondant].

In [15], we propose an unifying view of several recently proposed structured sparsity-inducing norms. We consider the situation of a model simultaneously (a) penalized by a set- function de ned on the support of the unknown parameter vector which represents prior knowledge on supports, and (b) regularized in Lp-norm. We show that the natural combinatorial optimization problems obtained may be relaxed into convex optimization problems and introduce a notion, the lower combinatorial envelope of a set-function, that characterizes the tightness of our relaxations. We moreover establish links with norms based on latent representations including the latent group Lasso and block-coding, and with norms obtained from submodular functions.

6.3. Kernel change-point detection

Participant: Sylvain Arlot [correspondant].

In [16], we tackle the change-point problem with data belonging to a general set. We propose a penalty for choosing the number of change-points in the kernel-based method of Harchaoui and Cappé (2007). This penalty generalizes the one proposed for one dimensional signals by Lebarbier (2005). We prove it satisfies a non-asymptotic oracle inequality by showing a new concentration result in Hilbert spaces. Experiments on synthetic and real data illustrate the accuracy of our method, showing it can detect changes in the whole distribution of data, even when the mean and variance are constant. Our algorithm can also deal with data of complex nature, such as the GIST descriptors which are commonly used for video temporal segmentation.

Collaboration with Alain Celisse (University Lille 1; Inria Lille, MODAL team) and Zaïd Harchaoui (Inria Grenoble, LEAR team).

6.4. On the Equivalence between Herding and Conditional Gradient Algorithms

Participants: Francis Bach [correspondant], Simon Lacoste-Julien, Guillaume Obozinski.

In [5], we show that the herding procedure of Welling (2009) takes exactly the form of a standard convex optimization algorithm–namely a conditional gradient algorithm minimizing a quadratic moment discrepancy. This link enables us to invoke convergence results from convex optimization and to consider faster alternatives for the task of approximating integrals in a reproducing kernel Hilbert space. We study the behavior of the different variants through numerical simulations. The experiments indicate that while we can improve over herding on the task of approximating integrals, the original herding algorithm tends to approach more often the maximum entropy distribution, shedding more light on the learning bias behind herding.

6.5. V-fold cross-validation and V-fold penalization in least-squares density estimation

Participant: Sylvain Arlot [correspondant].

In [22], we study V-fold cross-validation for model selection in least-squares density estimation. The goal is to provide theoretical grounds for choosing V in order to minimize the least-squares risk of the selected estimator. We first prove a non asymptotic oracle inequality for V-fold cross-validation and its bias-corrected version (V-fold penalization), with an upper bound decreasing as a function of V. In particular, this result implies V-fold penalization is asymptotically optimal. Then, we compute the variance of V-fold cross-validation and related criteria, as well as the variance of key quantities for model selection performances. We show these variances depend on V like $1 + 1/(V - 1)$ (at least in some particular cases), suggesting the performances increase much from $V = 2$ to $V = 5$ or 10, and then is almost constant. Overall, this explains the common advice to take $V = 10$—at least in our setting and when the computational power is limited—, as confirmed by some simulation experiments.

Collaboration with Matthieu Lerasle (CNRS, University Nice Sophia Antipolis).

6.6. Machine learning for Neuro-imaging

Participants: Fabian Pedregosa [correspondant], Francis Bach, Guillaume Obozinski.

In the course of the year 2011-2012 two articles where submitted and accepted in international workshops. The first published article, **Improved brain pattern recovery through ranking approaches** ([12]) was presented at the 2nd International Workshop on Pattern Recognition in NeuroImaging in London, July 2012 and proposes a new approach for the problem of estimating the coefficients of a generalized linear model with monotonicity constraint. For this, we explore the use of ranking techniques, which are popular in the context of information retrieval but novel for medical imaging applications.

The second published article, **Learning to rank from medical imaging data** ([11]) uses the same techniques as the previous article to solve a more fundamental problem, that is, to predict a quantitative (and potentially non-linear) variable from a set of noisy measurements. We show on simulations and two fMRI datasets that this approach is able to predict the correct ordering on pairs of images, yielding higher prediction accuracy than standard regression and multiclass classification techniques.

Collaboration with the Parietal project-team (A. Gramfort, B. Thirion, G. Varoquaux)

6.7. SiGMa: Simple Greedy Matching for Aligning Large Knowledge Bases

Participant: Simon Lacoste-Julien [correspondant].

The Internet has enabled the creation of a growing number of large-scale knowledge bases in a variety of domains containing complementary information. Tools for automatically aligning these knowledge bases would make it possible to unify many sources of structured knowledge and answer complex queries. However, the efficient alignment of large-scale knowledge bases still poses a considerable challenge. In [20], we present Simple Greedy Matching (SiGMa), a simple algorithm for aligning knowledge bases with millions of entities and facts. SiGMa is an iterative propagation algorithm which leverages both the structural information from the relationship graph as well as flexible similarity measures between entity properties in a greedy local search, thus making it scalable. Despite its greedy nature, our experiments indicate that SiGMa can efficiently match some of the world's largest knowledge bases with high precision. We provide additional experiments on benchmark datasets which demonstrate that SiGMa can outperform state-of-the-art approaches both in accuracy and efficiency.

Collaboration with Konstantina Palla, Alex Davies, Zoubin Ghahramani (Machine Learning Group, Department of Engineering, University of Cambridge); Gjergji Kasneci (Max Planck Institut fur Informatik); Thore Graepel (Microsoft Research Cambridge).

6.8. Block-Coordinate Frank-Wolfe Optimization for Structural SVMs

Participants: Simon Lacoste-Julien [correspondant], Mark Schmidt.

In [19], we propose a randomized block-coordinate variant of the classic Frank-Wolfe algorithm for convex optimization with block-separable constraints. Despite its lower iteration cost, we show that it achieves the same convergence rate in duality gap as the full Frank-Wolfe algorithm. We also show that, when applied to the dual structural support vector machine (SVM) objective, this yields an online algorithm that has the same low iteration complexity as primal stochastic subgradient methods. However, unlike stochastic subgradient methods, the stochastic Frank-Wolfe algorithm allows us to compute the optimal step-size and yields a computable duality gap guarantee. Our experiments indicate that this simple algorithm outperforms competing structural SVM solvers.

Collaboration with Martin Jaggi (Centre de Mathématiques Appliquées, Ecole Polytechnique); Patrick Pletscher (Machine Learning Laboratory, ETH Zurich).

6.9. A convex relaxation for weakly supervised classifiers

Participants: Armand Joulin [correspondant], Francis Bach.

In [8], we introduce a general multi-class approach to weakly supervised classification. Inferring the labels and learning the parameters of the model is usually done jointly through a block-coordinate descent algorithm such as expectation-maximization (EM), which may lead to local minima. To avoid this problem, we propose a cost function based on a convex relaxation of the soft-max loss. We then propose an algorithm specifically designed to efficiently solve the corresponding semidefinite program (SDP). Empirically, our method compares favorably to standard ones on different datasets for multiple instance learning and semi-supervised learning, as well as on clustering tasks.

6.10. Multi-Class Cosegmentation

Participants: Armand Joulin [correspondant], Francis Bach.

Bottom-up, fully unsupervised segmentation remains a daunting challenge for computer vision. In the cosegmentation context, on the other hand, the availability of multiple images assumed to contain instances of the same object classes provides a weak form of supervision that can be exploited by discriminative approaches. Unfortunately, most existing algorithms are limited to a very small number of images and/or object classes (typically two of each). In [9], we propose a novel energy-minimization approach to cosegmentation that can handle multiple classes and a significantly larger number of images. The proposed cost function combines spectral- and discriminative-clustering terms, and it admits a probabilistic interpretation. It is optimized using an efficient EM method, initialized using a convex quadratic approximation of the energy. Comparative experiments show that the proposed approach matches or improves the state of the art on several standard datasets.

Collaboration with the Willow project-team (J. Ponce).

6.11. A latent factor model for highly multi-relational data

Participants: Nicolas Le Roux, Guillaume Obozinski [correspondant].

Many data such as social networks, movie preferences or knowledge bases are multi-relational, in that they describe multiple relations between entities. While there is a large body of work focused on modeling these data, modeling these multiple types of relations jointly remains challenging. Further, existing approaches tend to breakdown when the number of these types grows. In [7], we propose a method for modeling large multi relational datasets, with possibly thousands of relations. Our model is based on a bilinear structure, which captures various orders of interaction of the data, and also shares sparse latent factors across different relations. We illustrate the performance of our approach on standard tensor-factorization datasets where we attain, or outperform, state-of-the-art results. Finally, a NLP application demonstrates our scalability and the ability of our model to learn efficient and semantically meaningful verb representations.

Collaboration with R. Jenatton (CMAP, Ecole Polytechnique) and Antoine Bordes (CNRS, Université de Technologie de Compiégne).

6.12. Semi-supervised NMF with time-frequency annotations for single-channel source separation

Participants: Francis Bach, Augustin Lefèvre [correspondant].

In [10], we formulate a novel extension of nonnegative matrix factorization (NMF) to take into account partial information on source-specific activity in the spectrogram. Results on single-channel source separation show that time-frequency annotations allow to disambiguate the source separation problem, and learned annotations open the way for a completely unsupervised learning procedure for source separation with no human intervention.

Collaboration with C. Févotte (Laboratoire traitement et communication de l'information (LTCI), CNRS: UMR5141 - Institut Télécom - Télécom ParisTech).

SIMPAF Project-Team

6. New Results

6.1. Interactions of Macro- and Microscopic scales

6.1.1. Homogenization methods

We have obtained three types of results regarding the homogenization theory and its applications. The first series of results is related to nonlinear elasticity. In [44], A. Gloria has proved the convergence of a discrete model for rubber towards a nonlinear elasticity theory in collaboration with R. Alicandro and M. Cicalese. This analysis has motivated the study of a specific random point set to model the stochastic network of polymer chains, namely the random parking measure, and results have been obtained by A. Gloria and M. Penrose (University of Bath) in [42]. The numerical simulation of the model with the random parking measure has been addressed by A. Gloria, P. La Tallec and M. Vidrascu (project team REO) in [21], and the comparisons with mechanical experiments are promising, A related inverse problem is currently under investigation by M. de Buhan, A. Gloria, P. Le Tallec, and M. Vidrascu.

A second type of results concerns a quantitative theory of stochastic homogenization of discrete linear elliptic equations. A breakthrough has been obtained by A. Gloria and F. Otto (MPI Leipzig) in [63] and [24], who gave the first optimal variance estimate of the energy density of the corrector field for stochastic discrete elliptic equations. The proof makes extensive use of a spectral gap estimate and of deep elliptic regularity theory, bringing in fact the probabilistic arguments to a minimum. This analysis has enabled A. Gloria to propose efficient numerical homogenization methods, both in the discrete and continuum settings [62], [20], see the review article [33]. In [23], A. Gloria and J.-C. Mourrat has pushed the approach forward and introduced new approximation formulas for the homogenized coefficient. In [22] they have considered a more probabilistic approach and given a complete error analysis of a Monte-Carlo approximation of the homogenized coefficients in the discrete case. Work in progress concerns the generalization of the results on discrete elliptic equations to the continuum case.

The third direction of research concerns the periodic homogenization of a coupled elliptic/parabolic system arising in the modelling of nuclear waste storage. This work is in collaboration with the French agency ANDRA. A. Gloria, T. Goudon, and S. Krell have made a complete theoretical analysis of the problem, derived effective equations, and devised an efficient method to solve the effective problem numerically, based on the reduced basis approach, see [41]. This subject has been pushed forward by Z. Habibi in collaboration with ANDRA.

6.1.2. Statistical physics : molecular dynamics

In [28], the analysis of constrained molecular dynamics is proposed, with associated numerical schemes.

In [29], the pobabilistic derivation of the chemotaxis equation from the individual motion of bacteriae have been carried out. In [30], a numerical method with asymptotic variance reduction have been proposed.

6.2. Plasmas

We investigated a projective integration scheme for a kinetic equation in the limit of vanishing mean free path, in which the kinetic description approaches a diffusion phenomenon. The scheme first takes a few small steps with a simple, explicit method, such as a spatial centered flux/forward Euler time integration, and subsequently projects the results forward in time over a large time step on the diffusion time scale. We showed that, with an appropriate choice of the inner step size, the time-step restriction on the outer time step is similar to the stability condition for the diffusion equation, whereas the required number of inner steps does not depend on the mean free path. We also provided a consistency result. The presented method is asymptotic-preserving, in the sense that the method converges to a standard finite volume scheme for the diffusion equation in the limit of vanishing mean free path. This is a joint work with G. Samaey (K. U. Leuven) [27].

6.3. Finite element and finite volume methods

6.3.1. Control in fluid mechanics

Recently, open and closed active flow control were carried out in order to study the flow behavior over a backward-facing step in a transitional regime. It was done either by a global frequency destabilization at the entry of the domain, or by a local blowing or suction through the lower and upper parts of the step by the use of small jets ([58], E. Creusé, A. Giovannini (IMFT Toulouse) and I. Mortazavi (MC2 Inria EPI, Bordeaux)). The numerical computations were based on a vortex-in-cell method. Such controls were shown to be efficient in reducing the average recirculation length value, the global flow energy, as well as the global flow enstrophy. We have now in mind to apply such a strategy on cavity-stent flows, in order to study the effect of passive and/or active control on the average emptying time of the cavity, corresponding to a lot of possible industrial or health applications (combustion, blood circulation in arteries,...).
Passive as well as active control were also performed on the "Ahmed body geometry", which can be considered as a first approximation of a vehicle profile. This work was carried out in collaboration with the EPI Inria MC2 team in Bordeaux (C.H. Bruneau, I. Mortazavi and D. Depeyras), as well as with Renault car industry (P. Gillieron). We recently combined active and passive control strategies in order to reach efficient results, especially concerning the drag coefficient, for two and three dimensional simulations [51]. We recently worked on a 25° rear-window configuration of the Ahmed body, for which the 3D-effects are very important and have to be considered in the numerical simulations [9]. Moreover, the effect of the vortices dynamics on the drag coefficient of a square Ahmed body was adressed [53], as well as the impact of several Ahmed bodies on the same road [52].

In another field of applications, a work was performed with the TEMPO Laboratory of Valenciennes. The objective of this collaboration was to study the pressure wave generated by high-speed trains entering tunnels in order to improve the shape of the tunnel sections.

6.3.2. Numerical Methods for viscous flows

In the case of compressible models, as the Euler equations, a careful analysis of sharp and practical stability conditions to ensure the positivity of both density and pressure variables was performed [11]. We are also concerned with the numerical simulation of certain multi-fluids flows, which in particular arises in the modelling of powdersnow avalanches. The hybrid scheme works on unstructured meshes and can be advantageously coupled to mesh refinements strategies in order to follow fronts of high density variation [38]. In order to answer these questions, we have developed a MATLAB code (NS2DDV-M, see the softwares section), a Fortran code and a C++ code.

6.3.3. A posteriori error estimators for finite element methods

A recent work, in collaboration with S. Nicaise (LAMAV, Valenciennes), was devoted to the derivation of some so-called "reconstruction estimators" based on gradient averaging, in order to provide lower and upper bounds of the error arising from a discontinuous Galerkin approximation of a diffusion problem [59].

At the same time, some equilibrated-type estimators were developed for the Reissner-Mindlin system arising in solid mechanics applications, for conforming and locking-free approximations, in the context of the PhD. of É. Verhille.

At last, a collaboration with the "Laboratoire d'électrotechnique et d'électronique de puissance de Lille (L2EP)" began two years ago, to derive a residual-based a posteriori error estimator for the Maxwell system in its vectorial and scalar potential formulation A/Φ (PhD of Z. Tang). The objective was to obtain a mathematical rigorous error indicator, in order to couple it with the automatic mesh generator used by EDF for very practical issues.

Some residual-type a posteriori error estimators were developed in the context of magnetostatic and magnetodynamic Maxwell equations, given in their potential and harmonic formulations. Here, the task was to found a relevant decomposition of the error in order to obtain the reliability of the estimator, with the use of ad-hoc interpolations. This work was realized in collaboration with the L2EP Laboratory (Laboratoire d'Electrotechnique et d'Electronique de Puissance de Lille, Lille 1 University), and gave rise to several contributions [17], [18], [32], [31], [65], obtained in the context of the Ph-D thesis of Zuqi Tang [2]. Then, other results about reconstructed a posteriori error estimators were obtained for Discontinuous Galerkin methods, applied to convection-reaction-diffusion equations [16].

6.4. Numerical anlaysis of Schrödinger equations

6.4.1. Absorbing boundary conditions

C. Besse continues his collaboration with X. Antoine (EPI Corida) and P. Klein. They construct in [3] some classes of absorbing boundary conditions for the two-dimensional Schrödinger equation with a time and space varying exterior potential and for general convex smooth boundaries. The construction is based on asymptotics of the inhomogeneous pseudodifferential operators defining the related Dirichlet-to-Neumann operator. Furthermore, a priori estimates are developed for the truncated problems with various increasing order boundary conditions. They propose in [34] some suitable discretization schemes of these ABCs and prove some semi-discrete stability results. Furthermore, the full numerical discretization of the corresponding initial boundary value problems is considered and simulations are provided to compare the accuracy of the different ABCs.

6.4.2. Semi-classical limit of the nonlinear Schrödinger equation

C. Besse works with R. Carles and F. Méhats (EPI Ipso). They consider in [36] the semiclassical limit for the nonlinear Schrödinger equation. They introduce a phase/amplitude representation given by a system similar to the hydrodynamical formulation, whose novelty consists in including some asymptotically vanishing viscosity. They prove that the system is always locally well-posed in a class of Sobolev spaces, and globally well-posed for a fixed positive Planck constant in the one-dimensional case. They propose a second order numerical scheme which is asymptotic preserving. Before singularities appear in the limiting Euler equation, they recover the quadratic physical observables as well as the wave function with mesh size and time step independent of the Planck constant. This approach is also well suited to the linear Schrödinger equation.

6.4.3. Analysis and numerical simulation of the Schrödinger equation

The linear or nonlinear Schrödinger equation with potential is one of the basic equations of quantum mechanics and it arises in many areas of physical and technological interest, e.g. in quantum semiconductors, in electromagnetic wave propagation, and in seismic migration. The Schrödinger equation is the lowest order one-way approximation (paraxial wave equation) to the Helmholtz equation and is called Fresnel equation in optics, or standard parabolic equation in underwater acoustics. The solution of the equation is defined on an unbounded domain. If one wants to solve such a whole space evolution problem numerically, one has to restrict the computational domain by introducing artificial boundary conditions. So, the objective is to approximate the exact solution of the whole-space problem, restricted to a finite computational domain. A review article [45] was written this year to describe and compare the different current approaches of constructing and discretizing the transparent boundary conditions in one and two dimensions. However, these approaches are limited to the linear case (or nonlinear with the classical cubic nonlinearity: an article written was dedicated to this case this year [49]) and constant potentials. Therefore, in collaboration with X. Antoine (IECN Nancy and Inria Lorraine), we proposed to P. Klein to study, in her PhD thesis, the case of the Schrödinger equation with variable potentials. The study of the non-stationary one-dimensional case has already led to one publication [46] and some preliminary results in the stationary case are really promising. These cases are relevant since for example the equations appear in the Bose Einstein condensate with a quadratic potential.

This problem is obviously not limited to the Schrödinger equation and new developments are in progress on the Korteweg de Vries equation with M. Ehrhardt. This equation is more difficult to study due to its third order derivative in space.

Dispersive equations, such as the Schrödinger equation are also considered as boundary-value problems. For example, in [60], G. Dujardin studies the long time asymptotics of the solutions of linear Schrodinger equations considered as initial-boundary value problems on the half-line and on bounded intervals when the boundary data are periodic functions of time. G. Dujardin obtains theoretical results using a transformation method introduced by T. Fokas and provides several numerical experiments to support them.

6.5. Other contributions

6.5.1. Corrosion models

The Diffusion Poisson Coupled Model [47] is a model of iron based alloy in a nuclear waste repository. It describes the growth of an oxide layer in this framework. The system is made of a Poisson equation on the electrostatic potential and convection-diffusion equations on the densities f charge carriers (electrons, ferric cations and oxygen vacancies), supplemented with coupled Robin boundary conditions. The DPCM model also takes into account the growth of the oxide host lattice and its dissolution, leading to moving boundary equations. In [12], C. Chainais-Hillairet and I. Lacroix-Violet consider a simplified version of this model, where only two charge carriers are taken into account and where there is no evolution of the layer thickness. They prove the existence of a steady-state solution to this model. More recently, C. Chainais-Hillairet and I. Lacroix-Violet have also obtained an existence result for the time-dependent simplified model. This result will be soon submitted for publication.

In [4], C. Chainais-Hillairet and coworkers have studied some numerical methods for the approximation of the DPCM model. The choice of the numerical methods is justified by a stability analysis and by the study of their numerical performance. These methods have been implemented in the code CALIPSO developed at ANDRA. Numerical experiments with real-life data show the efficiency of the developed methods.

6.5.2. Transparent boundary conditions for the Helmholtz equation

C. Besse and I. Violet start a collaboration with S. Fliss (Poems), J. Coatleven and K. Ramdani (Corida) to build transparent boundary conditions for the Helmholtz equation. They propose in [6] a strategy to determine the Dirichlet-to-Neumann (DtN) operator for infinite, lossy and locally perturbed hexagonal periodic media. They obtain a factorization of this operator involving two non local operators. The first one is a DtN type operator and corresponds to a half-space problem. The second one is a Dirichlet-to-Dirichlet (DtD) type operator related to the symmetry properties of the problem. The half-space DtN operator is characterized via Floquet-Bloch transform, a family of elementary strip problems and a family of stationary Riccati equations. The DtD operator is the solution of an affine operator valued equation which can be reformulated as a non standard integral equation.

6.5.3. Analysis of subcycling techniques

Several physics situations involve phenomena which occur on very different time scales. A popular option to integrate the equations in time in this context is to use sub-cycling techniques, which allow to weaken the stability constraints. Several questions are still open for the asymptotic behavior of such methods, *e.g.* the preservation of equilibrium states. New results about the asymptotic orders if such methods have been derived on toy-model problems which allow a better understanding of these methods and their preservation of equilibrium states [40].

6.5.4. Phase transitions

We analyzed numerically a forward-backward diffusion equation with a cubic- like diffusion function, –emerging in the framework of phase transitions modelling– and its “entropy” formulation determined by considering it as the singular limit of a third-order pseudo-parabolic equation. Precisely, we proposed schemes for both the second and the third order equations, we discussed the analytical properties of their semi-discrete counter- parts and we compared the numerical results in the case of initial data of Riemann type, showing strengths and flaws of the two approaches, the main emphasis being put on the propagation of transition interfaces. This is a joint work with C. Mascia (Univ. La Sapienza) [25].

6.5.5. Modelling of the biological populations

We worked on two problems of biological populations: the understanding of the occurrence of collective behavior for large populations and the extinction probabilities in some population dynamics.

Several approaches are used in the modelling of collective behavior models for large populations of fish : we obtained results at the particle and kinetic levels for a model involving self-propulsion, friction and an attractive/repulsive potential. By introducing a new dimensionless setting, we identified five parameters that govern the possible asymptotic states for this system (clumps, spheres, dispersion, mills, rigid-body rotation, flocks) and performed a numerical analysis on the 3D particle-setting. Also, we described the kinetic system derived as the limit from the particle model as N tends to infinity; and we proposed, in 1D, a numerical scheme for the simulations, and performed a numerical analysis devoted to trying to recover asymptotically patterns similar to those emerging for the equivalent particle systems, when particles originally evolved on a circle. this is a joint work with J. Rosado (UCLA) and F. Vecil (Univ. Valencia) [43].

The extinction probabilities of a flower population may be modelled by an imhomogeneous random walk on the positive quadrant. On the one hand, introducing the generating function, that solves a PDE, we computed an explicit solution. On the other hand, we compared stochastic and deterministic resolutions of the random walk. This is a joint work with K. Raschel (Univ. Tours), V. C. Tran (Univ. Lille 1) [26].

TAO Project-Team

6. New Results

6.1. Realistic step sizes for optimization algorithms

Many theoretical results about objective improvement in the process of continuous optimization rely on the assumption that the steps of the algorithm are infinitesimally small, the only situation in which theoretical guarantees of improvement can be given. Y. Akimoto and Y. Ollivier have waived the necessity for such an assumption in a whole class of continuous optimization algorithms, thanks to the use of information geometry [20]. This takes theory closer to the practice of actual optimization algorithms.

6.2. Noisy Optimization Bounds with Constant Noise Variance

Many bounds in noisy evolutionary optimization are based on low variance assumptions (in particular, variance of noise converging to 0 close to the optima). Other bounds in the optimization literature consider difficult objective functions. We prove some new bounds, in the following setting[55]:

- without assuming that the variance is going to zero at the optimum;
- following some debates on the COCO mailing list (see 5.4), assuming that sampling far from the optimum (we had earlier results without this assumption; new results emphasize the contrast).

6.3. Extensions of Upper Confidence Trees

We developed extensions of Upper Confidence Trees to continuous or large domains (states and/or actions) and to domains with high expertise or strong structure[37], [31], [38] (incidentally realizing performances on MineSweeper); we recently submitted a proof of a variant of UCT with consistency proof in the continuous domains (both actions and random variables are allowed to be continuous). Another extension is to the difficult setting with no possibility to “undo” a decision or duplicate a state; see [63]. Yet another extension aims at multi-objective optimization [56].

6.4. Mixing myopic fast algorithms and asymptotically optimal algorithms

We made several works based on combining in sequential decision making:

- a fast algorithm providing quickly good heuristic results;
- an asymptotically optimal, too slow for real size problems.

Results are published in [31] and [38], outperforming the state of the art for MineSweeper in reasonable time; an application to energy has been done, and a new one is under work (see Section 4.1). We believe that this diea of combining fast approximate solutions and slow asymptotically optimal algorithms is a key for improving the state of the art in high dimensional combinatorial planning and that our results on MineSweeper and moderate size energy problem are a solid first step in this direction.

6.5. Adaptive Metropolis with Online Relabeling

In [23] we proposed a novel adaptive MCMC algorithm named AMOR (Adaptive Metropolis with Online Relabeling) for efficiently simulating from permutation-invariant targets occurring in, for example, Bayesian analysis of mixture models. An important feature of the algorithm is to tie the adaptation of the proposal distribution to the choice of a particular restriction of the target to a domain where label switching cannot occur. The algorithm relies on a stochastic approximation procedure for which we design a Lyapunov function that formally defines the criterion used for selecting the relabeling rule. This criterion reveals an interesting connection with the problem of optimal quantifier design in vector quantization which was only implicit in previous works on the label switching problem. In benchmark examples, the algorithm turns out to be fast-converging and efficient at selecting meaningful non-trivial relabeling rules to allow accurate parameter inference. In [24] the algorithm was applied to a synthetic mixture model inspired by the muonic water Cherenkov signal of the surface detectors in the Pierre Auger Experiment.

6.6. Reinforcement learning for frugal cascade learning

In [32] we propose an algorithm that builds sparse decision DAGs (directed acyclic graphs) from a list of base classifiers provided by an external learning method such as AdaBoost. The basic idea is to cast the DAG design task as a Markov decision process. Each instance can decide to use or to skip each base classifier, based on the current state of the classifier being built. The result is a sparse decision DAG where the base classifiers are selected in a data-dependent way. The method has a single hyperparameter with a clear semantics of controlling the accuracy/speed trade-off. The algorithm is competitive with state-of-the-art cascade detectors on three object-detection benchmarks, and it clearly outperforms them when there is a small number of base classifiers. Unlike cascades, it is also readily applicable for multi-class classification. Using the multi-class setup, we show on a benchmark web page ranking data set that we can significantly improve the decision speed without harming the performance of the ranker. Beside outperforming classical cascade designs on benchmark data sets, the algorithm also produces interesting deep structures where similar input data follows the same path in the DAG, and subpaths of increasing length represent features of increasing complexity.

TOSCA Project-Team

6. New Results

6.1. Probabilistic numerical methods, stochastic modelling and applications

Participants: Mireille Bossy, Nicolas Champagnat, Julia Charrier, Julien Claisse, Madalina Deaconu, Samuel Herrmann, James Inglis, Antoine Lejay, Sylvain Maire, Sebastian Niklitschek Soto, Nicolas Perrin, Denis Talay, Etienne Tanré, Denis Villemonais, Laurent Violeau.

6.1.1. Published works and preprints

- In collaboration with P.-E. Jabin (University of Maryland), J.-F. Jabir and J. Fontbona (CMM and Universidad de Chile, Santiago de Chile), M. Bossy have studied the link between the Lagrangian version of divergence free constraint (and the uniform density constraint), with an additional potential term in the Lagrangian equation, having some similarity with the role of the Eulerian pressure term. They obtained the local existence of analytical solutions to an incompressible Lagrangian stochastic model in periodic domain. The paper is in positive revision for publication in *Communications in Partial Differential Equations* [33]. http://hal.inria.fr/hal-00691712
- N. Champagnat worked with A. Lambert (Univ. Paris 6) on splitting trees with Poissonian mutations. Assuming that each mutation is neutral and gives a new type in the population, they obtained in [13], [14] large time convergence results on the sizes of the largest families and the ages of the oldest families in the population. http://hal.inria.fr/inria-00515481, http://hal.inria.fr/inria-00616765. In collaboration with Mathieu Richard (Ecole Polytechnique, Palaiseau), they also extended some of these results to the case of splitting trees with mutations occuring at birth of individuals [15], http://hal.inria.fr/hal-00736036.
- N. Champagnat obtained with P. Diaconis (Stanford Univ.) and L. Miclo (Univ. Toulouse 3) the full spectral decomposition of the transition matrix of two-dimensional Markov chains $(X_n, Y_n)_{n\geq 0}$ in $\mathbb{Z}_+^2$, without immigration or mutation, which are *neutral* in the sense that $(X_n+Y_n)_{n\geq 0}$ is a Markov process. Because of the specific form of the eigenvectors, they were also able to characterize all the Dirichlet eigenvectors in subdomains of $\mathbb{Z}_+^2$ of the form $\{(i,j)\in\mathbb{Z}_+^2 : i+j\geq d\}$ for all $d\geq 0$. As an application, they could determine the quasi-stationary and quasi-limiting distributions of such processes [12], http://hal.inria.fr/hal-00672938.
- N. Champagnat studied with F. Campillo (EPI Modemic, Inria Sophia Antipolis — Méditerrannée) individual based models of clonal plants where plants interact through the network formed by the rizhomes or stolons linking plants. In the limit of large population, they obtained a PDE governing the dynamics of population densities in space [11], http://hal.inria.fr/hal-00723209.
- M. Deaconu and S. Herrmann introduced a new method for the simulation of the hitting times of nonlinear boundaries for Bessel processes. This method combines the method of images and the random walk on spheres method. They construct the so called walk on moving spheres algorithm. This approach can be applied for the hitting time of a given level for the Cox-Ingersoll-Ross process and thus be used in models coming from finance and neuroscience [17], http://hal.inria.fr/hal-00636056/en. This work is part of the ANR MANDy project.
- J. Inglis and E. Tanré studied with F. Delarue and S. Rubenthaler (Univ. Nice – Sophia Antipolis) the global solvability of a networked system of integrate-and-fire neurons proposed in the neuroscience literature. In the mean-field limit the equation resembles a McKean-Vlasov equation, but is highly non-standard and previous attempts at rigorous analysis were not satisfactory. They here bridge this gap, and shed light on a surprisingly complicated problem [35], http://hal.inria.fr/hal-00747565.

- A. Lejay continued his long term investigations on probabilistic interpretations and Monte Carlo simulations of interfaces conditions, such as ones arising in discontinuous media. With G. Pichot (IRISA, Rennes), he has developed a series of tests and benchmarks regarding one-dimensional Monte Carlo methods, such as the ones proposed in [19], http://hal.inria.fr/hal-00649170. He has also developed a new family of stochastic diffusion processes, called the *snapping out Brownian motion*, in order to take into account an interface condition where the concentration of the fluid is proportional to its gradient. Finally, A. Lejay and S. Maire also proposed new methods and tested a few ones to deal with the locally isotropic case for multidimensional problems [18], http://hal.inria.fr/hal-00689581.
- With A. Kohatsu-Higa (Ristumeikan University) and K. Yasuda (Hosei university), A. Lejay has continued his work [25] on the simulation of SDE with a discontinuous drift. http://hal.inria.fr/hal-00670123
- With L. Coutin (University of Toulouse), A. Lejay has developed an appropriate framework to deal with linear rough differential equations, extending some results (Magnus formula, Dyson series...) to this case. Using theses properties, they have studied the sensitivites of solutions of rough differential equations with respect to the signal, the vector field or the starting point. They have provided new results such as the Hölder continuity of the derivative of the so called Itô map which transforms a rough path to the solution of a rough differential equation [34]. http://hal.inria.fr/hal-00722900
- S. Maire and C. Prissette (Univ. du Sud – Toulon – Var) have developed in [21] a stochastic algorithm to solve Sudoku puzzles using estimation of distribution coupled with restart techniques. http://hal.inria.fr/inria-00591852
- S. Maire and E. Tanré have generalised the spectral methods for elliptic PDEs developed in [42], [43] to the case of pure Neumann boundary conditions. Some additional difficulties occur because the stochastic representation of the solutions is defined only up to an additive constant and as a limit involving local time approximations [40]. By taking into account these additional properties, they still obtained a spectral matrix having a condition number converging to one [36]. http://hal.inria.fr/hal-00677529
- C. Graham (Ecole Polytechnique) and D. Talay wrote the first volume [27] of their series of books published by Springer on the Foundations of Stochastic Simulations. They started to write the second volume.
- D. Villemonais wrote with S. Méléard (École Polytechnique) a survey on quasi-stationary distributions and Q-processes for stochastic models of population dynamics. This survey also contains a detailed numerical study of the behaviour of classical models with extinction [23]. http://hal.inria.fr/hal-00653834
- D. Villemonais worked on the empirical distribution of Fleming-Viot type particle systems. Using couplings with reflected diffusion processes, he proved the uniform tightness of such empirical distributions and deduced the non-degeneracy of the law of diffusion processes conditioned not to hit a boundary [39]. http://hal.inria.fr/hal-00681601
- D. Villemonais proved in [38] a general approximation method for Markov processes conditioned not be killed. The method is based on a mean field interacting particles system which is easy to simulate. The study also details the particular case of time/environment dependent diffusion processes. http://hal.archives-ouvertes.fr/hal-00598085

6.1.2. Other works in progress

- N. Champagnat and D. Villemonais obtained criterions for existence and uniqueness of quasi-stationary distributions and Q-processes for general absorbed Markov processes. A quasi-stationary distribution is a stationary distribution conditionnally on non-absorbtion, and the Q-process is defined as the original Markov process conditionned to never be absorbed. The criterion that they obtain also ensures exponential convergence of the conditionned t-marginal of the process conditionned not to be absorbed at time t to the quasi-stationary distribution and the exponential ergodicity of the Q-process. This work is currently being written.

- N. Champagnat and D. Villemonais work on time-reversal of absorbed processes, which allow to characterize the path to extinction in extinct populations which are known to be non-extinct at some time in the past. They plan to apply these results on practical ecological situations.
- J. Claisse continued his PhD. under the supervision of N. Champagnat and D. Talay on stochastic control of population dynamics. He completed a finite-horizon and an infinite-horizon optimal control problem on a birth-death process. He is currently working on a finite-horizon optimal control problem on a branching-diffusion process. In addition, he is working on modelling of a pH-mediated cancer treament.
- M. Deaconu and S. Herrmann continue the study of the hitting times for Bessel processes in the situation of noninteger dimensions and also in the application of this method to the simulation of the Brownian hitting time,
- M. Deaconu starts a collaboration with L. Beznea (Simion Stoilow Institute of Mathematics of the Romanian Academy) on coagulation-fragmentation models and their connection with branching processes.
- M. Deaconu studies in collaboration with F. Nobile and F. Tesei (EPFL) a pollution model by using hitting times of stochastic processes.
- S. Herrmann and E. Tanré worked on a scheme to construct an efficient algorithm to simulate the first hitting time of curves by a one dimensional Brownian motion. They apply the result to estimate the spiking time of leaky integrate fire models in neuroscience. This work is part of the ANR MANDy project.
- S. Larnier joined the team in September as a post-doctoral researcher and began working with A. Lejay on data assimilation in order to predict the ocean wave energy from the knowledge of near-shore incoming waves. They started a collaboration on video data with R. Almar (LEGOS, Toulouse) and R. Cienfuegos (Pontificia Universidad Católica de Chile).
- S. Maire works with M. Simon (Mainz Univ.) on electrical impedance tomography problems using new Monte Carlo schemes that deal with Robin and transmission boundary conditions.
- S. Maire develops with I. Dimov (Bulgarian academy of sciences) a Monte Carlo method called the walk on equations to solve linear systems of algebraic equations.
- S. Niklitschek has continued his PhD. work under the supervision of D. Talay. They were able to extended their first work in which they gave a probabilistic interpretation of a parabolic equation with discontinuous drift and proved the weak rate of convergence of the Euler method using the accurate pointwise estimates obtained for the derivatives of the solution, to the case in which both drift and diffusion coefficients are discontinuous. Both results are consistent with each other, and also with the results obtained by M. Martinez and D. Talay in [22].
- N. Perrin continued his PhD. on stochastic methods in molecular dynamics under the supervision of M. Bossy, N. Champagnat and D. Talay. This year, he studied a stochastic interpretation of parabolic PDEs with divergence form operators involved in the Poisson-Boltzmann PDE of molecular dynamics, and the associated numerical Monte Carlo method. He also continued his study of a method due to P. Malliavin (French Academy of Science) based on the Fourier analysis of covariance matrices with delay in order to identify the fast and slow components of a molecular dynamics.
- P. Guiraud (University of Valparaiso) and E. Tanré study the effect of noise in the phenomenon of spontaneous synchronisation in a network of full connected integrate-and-fire neurons. They detail cases in which the phenomenon of synchronization persists in a noisy environment, cases in which noise permits to accelerate synchronization, and cases in which noise permits to observe synchronization while noiseless model does not have synchronization.
- P. Orio (Centro Interdisciplinario de Neurociencia de Valparaiso) and E. Tanré work on the comparison of global properties of the solution of mathematical models and the associated measurements obtained by experiments.

- L. Violeau continued his PhD. on *Stochastic Lagrangian Models and Applications to Downscaling in Fluid Dynamics* under the supervision of M. Bossy and A. Rousseau (MOISE team, Inria Sophia Antipolis – Méditerranée, Montpellier). Laurent studied this year the rate of convergence of the Nadaraya-Watson conditional estimator for "linear" kinetic processes. He is currently working on the rate of convergence of the particle approximation of kinetic conditional McKean-Vlasov stochastic models.
- P-E. Jabin and D. Talay continue to develop their innovating approach, which combines stochastic analysis and PDE analysis, for the time varying Hamilton-Jacobi-Bellman-McKean-Vlasov equations of the Lasry and Lions mean-field stochastic control theory.
- D. Talay is working with J. Bion-Nadal (Ecole Polytechnique) on applications of risk measures to the calibration of stochastic models, with N. Touzi (Ecole Polytechnique) on the stochastic control of stochastic differential equations with weighted local times, and with O. Bardou (GDF) on Edgeworth expansions for the Central Limit Theorem for Brownian martingales whose integrands depend on ergodic diffusion processes.

6.2. Financial Mathematics

Participants: Mireille Bossy, Paul Charton, Dalia Ibrahim, Denis Talay, Etienne Tanré.

- Mireille Bossy, in collaboration with H. Quinteros (Univ. Chile) worked on the rate of convergence of non Lipschitz diffusion processes discretized with the symetrized Milstein scheme. Under the same kind of hypotheses than in [41] on the symetrized Euler scheme, they obtained the expected improvement of the strong rate of convergence, when the diffusion coefficient is of the form $\sigma(x) = x^\alpha$, with $\alpha \in [1/2, 1[$.

  A preprint is being written.
- P. Charton continued his PhD. under the supervision of M. Deaconu and A. Lejay. He studied some storage strategies for wind farms.
- **Mathematical modelling for technical analysis techniques** Since November 2009, D. Ibrahim has been working on her PhD. thesis on Mathematical modeling of technical analysis in finance, under supervision of D. Talay and E. Tanré. The aim of her work is to study the performances of a technical analysis tool designed to detect changes in the volatility term: The Bollinger Bands. She studied the performances of this indicator in a modified Black-Scholes model such that the volatility is equal to σ_0 up to a random time τ, independent of the Brownian motion governing the prices. After τ, the volatility is equal to σ_1. She proved that Bollinger Bandwidth indicator can detect the time change (at which the volatility changes its value), in the case of small and large volatilities. She has also exhibited a mathematical optimal allocation strategy, by decomposing the initial allocation problem into an allocation problem before the change time τ and an allocation problem after τ, in order to circumvent some technical problems brought from the change of volatility.

  This work is part of the contract with FINRISK.
- In collaboration with C. Michel (CA-CIB) and V. Reutenauer (Citi), D. Talay and E. Tanré worked on the
  - the study of the liquidity risk in the interest rate options market;
  - the minimization of the hedging error in interest rates Gaussian models by means of strategies designed in an effective way by using stochastic optimization algorithms.
- P. Protter (Columbia University) and D. Talay continue to work on bubbles time evolution models, which leads them to try to extend Feller's results on explosion times for stochastic differetential equations.

ABS Project-Team

5. New Results

5.1. Modeling Interfaces and Contacts

Docking, scoring, interfaces, protein complexes, scoring functions, Voronoi diagrams, arrangements of balls.

5.1.1. Modeling Macro-molecular Complexes : a Journey Across Scales

Participants: Frédéric Cazals, Tom Dreyfus.

In collaboration with C. Robert (IBPC / CNRS, Paris, France).

While proteins and nucleic acids are the fundamental components of an organism, Biology itself is based on the interactions they make with each other. Analyzing macromolecular interactions typically requires handling systems involving from two to hundreds of polypeptide chains. After a brief overview of the modeling challenges faced in computational structural biology, the text [16] reviews concepts and tools aiming at improving our understanding of the link between the static structures of macromolecular complexes and their biophysical/biological properties. We discuss geometrical approaches suited to atomic-resolution complexes and to large protein assemblies; for each, we also present examples of their successful application in quantifying and interpreting biological data. This methodology includes state-of-the-art geometric analyses of surface area, volume, curvature, and topological properties (isolated components, cavities, voids, cycles) related to Voronoi constructions in the context of structure analysis. On the applied side, we present novel insights into real biological problems gained thanks to these modeling tools.

5.1.2. CSA: Comprehensive Comparison of Pairwise Protein Structure Alignments

Participant: Noël Malod-Dognin.

In collaboration with I. Wohlers (CWI / VU University Amsterdam, Netherlands), R. Andonov (Irisa / Rennes University, France), G.W. Klau (CWI / VU University Amsterdam, Netherlands).

Protein structural alignment is a key method for answering many biological questions involving the transfer of information from well-studied proteins to less well-known proteins. Since structures are more conserved during evolution than sequences, structural alignment allows for the most precise mapping of equivalent residues. Many structure-based scoring schemes have been proposed and there is no consensus on which scoring is the best. Comparative studies also show that alignments produced by different methods can differ considerably. Based on the alignment engine derived from A_purva, we designed CSA (Comparative Structural Alignment), the first web server for computation, evaluation and comprehensive comparison of pairwise protein structure alignments at single residue level [15]. It offers the exact computation of alignments using the scoring schemes of DALI, Contact Map Overlap (CMO), MATRAS and PAUL. In CSA, computed or uploaded alignments can be explored in terms of many inter-residue distances, RMSD, and sequence-based scores. Intuitive visualizations also help in grasping the agreements and differences between alignments. The user can thus make educated decisions about the structural similarity of two proteins and, if necessary, post-process alignments by hand. CSA is available at http://csa.project.cwi.nl.

Upon publication [15], CSA was selected by *Nucleic Acids Research* as featured article of July 2012 (top 5% of papers in terms of originality, significance and scientific excellence.

5.2. Modeling Macro-molecular Assemblies

Macro-molecular assembly, reconstruction by data integration, proteomics, modeling with uncertainties, curved Voronoi diagrams, topological persistence.

5.2.1. Stoichiometry Determination for Mass-spectrometry Data: the Interval Case

Participants: Deepesh Agarwal, Frédéric Cazals, Noël Malod-Dognin.

In structural proteomics, given the individual masses of a set of protein types and the exact mass of a protein complex, the *exact stoichiometry determination problem (SD)*, also known as the money-change problem, consists of enumerating all the stoichiometries of these types which allow to recover the target mass. If the target mass suffers from experimental uncertainties, the *interval SD problem* consists of finding all the stoichiometry vectors compatible with a target mass within an interval.

We make contributions in two directions [18]. From a theoretical standpoint, we present a constant-memory space algorithm (`DIOPHANTINE`) and an output sensitive dynamic programming based algorithm (`DP++`), both inherently addressing the interval SD problem. From an applied perspective, we raise three points. First, we show that `DIOPHANTINE` and `DP++` yield an improvement from 3 to 4 orders of magnitude over state-of-the-art exact SD algorithms, for typical protein complexes facing uncertainties on the target mass in the range 0.1-1%. Second, we show that `DIOPHANTINE` behaves like an output-sensitive algorithm—especially when the interval width increases, albeit such a property cannot be expected in general. Third, from a biological perspective, using a panel of biological complexes (eukaryotic translation factor, yeast exosome, 19S proteasome sub-unit, nuclear pore complex), we stress the importance of enumeration, even at a null noise level.

The programs accompanying this paper are available from http://team.inria.fr/abs/addict/.

5.3. Algorithmic Foundations

Voronoi diagrams, α-shapes,

The work undertaken in this vein in 2012 will be finalized in 2013.

5.4. Immunology

Immune response, infection, antibodies, complementarity determining region (CDR)

5.4.1. Teleost Fish Mount Complex Clonal IgM and IgT Responses in Spleen Upon Systemic Viral Infection

Participant: Frédéric Cazals.

In collaboration with

- *R. Castro, L. Journeau, A. Benmansour and P. Boudinot (INRA Jouy-en-Josas, France)*
- *H.P. Pham and A. Six (Univ. of Paris VI, France)*
- *O. Bouchez (INRA Castanet Tolosan, France)*
- *V. Giudicelli and M-P. Lefranc (IMGT / CNRS, Montpellier, France)*
- *E. Quillet (INRA Jouy-en-Josas, France)*
- *S. Fillatreau (Leibniz Institute, Berlin, Germany)*
- *O. Sunyer (Univ. of Pennsylvania, USA)*

Upon infection, B-lymphocytes expressing antibodies specific for the intruding pathogen develop clonal responses triggered by pathogen recognition via the B-cell receptor. The constant region of antibodies produced by such developing clones dictates their functional properties. In teleost fish, the clonal structure of B-cell responses and the respective contribution of the three isotypes IgM, IgD, and IdT remains unknown. The expression of IgM and IgT are mutually exclusive, leading to the existence of two B-cell subsets expressing either both IgM and IgD or only IgT. In [12], we undertook a comprehensive analysis of the variable heavy chain (VH) domain repertoires of the IgM, IgD and IgT in spleen of homozygous isogenic rainbow trout (Onchorhynus mykiss), before and after challenge with a rhabdovirus, the Viral Hemorrhagic Septicemia Virus (VHSV), using CDR3-length spectratyping and pyrosequencing of immunoglobulin (Ig) transcripts. In healthy fish, we observed distinct repertories for IgM, IgD and IgT respectively, with a few amplified μ

and τ junctions, suggesting the presence of IgM and IgT secreting cells in the spleen. In infected animals, we detected complex and highly diverse IgM responses involving all VH subgroups, and dominated by a few large public and private clones. A lower number of robust clonal responses involving only a few VH were detected for the mucosal IgT, indicating that both IgM^+ and IgT^+ spleen B cells responded to systemic infection but at different degrees. In contrast, the IgD response to the infection was faint. Although the IgD and IgT present different structural features and evolutionary origin compared to mammalian IgD and IgA respectively, their implication in the B-cell response evokes these mouse and human counterparts. Thus, it appears that the general properties of antibody responses were already in place in common ancestors of fish and mammals, and were globally conserved during evolution with possible functional convergences.

AMIB Project-Team

5. New Results

5.1. RNA structures

5.1.1. RNA structure alignment

It is widely accepted that, for a large number of RNA families, the structure is more conserved than the sequence. Therefore, any reasonable notion of homology should consider the similarity in the secondary structure, i.e. how well the base-pairing positions in two structures can be put in correspondence, or aligned. In collaboration with a significant part of the French bioinformatics community, an assessment of the quality of existing algorithms for the problem was proposed [6]. Furthermore, a review of the state-of-the-art in RNA comparison algorithms is to be published [11], and a chapter in a forthcoming book on RNA computational biology was written in collaboration with Robert Giegerich (University Bielefeld) during his stay.

Most existing alignment tools rely on the assumption that the RNA structure is free of pseudoknots, i.e. free of crossing interactions. This condition naturally arises from the intractability of the unconstrained version of the problem. In a joint work, A. Denise, Ph. Rinaudo and Y. Ponty worked around this issue by proposing a parameterized complexity algorithmic solution for the unconstrained version of the problem. One of the key feature of this algorithm is that, although exponential in the worst-case scenario, it naturally adapts its complexity to the level of intricacy of the aligned structures, and remains polynomial for large classes of pseudoknots. Preliminary results of this work were presented at the WABI'12 conference [35].

5.1.2. Energy-weighted RNA algorithmics

We complemented previous studies led within AMIB on RNA structures with restricted classes of pseudoknots by showing, in a collaboration with Rolf Backofen (Freiburg University), that the computational hardness of RNA folding with general pseudoknots is extremely robust to the choice of a precise energy model. It was shown that the problem is completely unapproximable when expressive – yet realistic – energy models are taken into consideration. These results were presented at CPM'12 [37] (Helsinki, Finland).

Moreover, using an interpolation technique introduced at the RECOMB'11 conference, we were able to improve both the sequential and parallel complexities of the `RNAbor` algorithm developed within P. Clote's lab. The resulting algorithm and its application to the detection of conformational switches in sequence lengths that were previously unreachable by the algorithm, are described in a manuscript accepted in *Plos One*.

5.1.3. RNA knowledge-based potentials and 3D studies

The building of an RNA potential proved much harder and interesting than we initially expected. A non-redundant dataset had first to be extracted from the literature as the available dataset were not suitable for our study even the very recent ones. From the collected distance data, the building of a knowledge-based potential was usually done using histograms; and the histogram interval size and data fitting was an issue. In our 2012 study, we showed that the best solution to build potentials with no interval issue is by using Dirichlet Process Mixture Models (DPMs) [24]. We also benefited of the group experience in modeling the dynamics of RNA and normal-mode experiments to obtain two good decoy sets which complemented the well-known Farna study. We also showed that in many case our high-resolution predictions were better than the Farna/Rosetta standard.

5.1.4. RNA 3D structure prediction

In collaboration with PRISM at Versailles and Westhof's group at Strasbourg, we addressed the problem of ab initio prediction of RNA three-dimensional structure. We developed an algorithm for automatically predicting the topological family of any RNA three-way junction, an thus its coarse-grained local geometry, given only the information from the secondary structure: the sequence and the Watson–Crick pairings. Additionally, we showed that the results are noticeably improved if homology information is used [14]. The resulting software, `Cartaj`, is available online and downloadable at http://cartaj.lri.fr. Then we investigated a new approach for the global prediction of the coarse-grain 3D structure of RNA molecules. We model a molecule as being made of helices and junctions. Using our results above, we are able to classify junctions into topological families that determine their preferred 3D shapes. All the parts of the molecule are then allowed to establish long-distance contacts that induce a threedimensional folding of the molecule. An algorithm relying on game-theory was proposed to discover such long-distance contacts that allow the molecule to reach a Nash equilibrium. As reported by our experiments, this approach allows one to predict the global shape of large molecules of several hundreds of nucleotides that are out of reach of the state-of-the-art methods [15].

A graph-theoretic approach has been successfully used for classification and structure prediction of transmembrane beta-barrel proteins[23], [25].

5.2. Proteins structures and interactions

5.2.1. Protein-protein interaction

Adrien Guilhot, PhD candidate in our project worked on a modified scoring function for the Rosetta software suite. After an extensive conformation generation for the two recently published benchmarks, we now have a model for protein-RNA semi-flexible docking which is currently being tested.

The prediction of the network of protein-protein interactions (PPI) of an organism is crucial for the understanding of biological processes and for the development of new drugs. Machine learning methods have been successfully applied to the prediction of PPI in yeast by the integration of multiple direct and indirect biological data sources. However, experimental data are not available for most organisms. We propose in [9] an ensemble machine learning approach for the prediction of PPI that depends solely on features independent from experimental data. New estimators of the coevolution between proteins have been developed and combined them in an ensemble learning procedure.

This method has been applied to a dataset of known co-complexed proteins in *Escherichia coli* and compared it to previously published methods. Our method allows prediction of PPI with an unprecedented precision of 95.5% for the first 200 sorted pairs of proteins compared to 28.5% on the same dataset with the previous best method.

A close inspection of the best predicted pairs allowed us to detect new or recently discovered interactions between chemotactic components, the flagellar apparatus and RNA polymerase complexes in *E. coli*.

5.3. Combinatorics and Annotation

5.3.1. Word counting and random generation

A long-term research on word enumeration has been realized by the team, in order to calculate a statistical significance for a pattern occurrence according to a given background model. As a part of E. Furletova's thesis, defended in February 2012, co-advised by M. Roytberg (IMPB, Puschino, Russia) and M. Régnier, an extension to Hidden Markov Models, `SufPref`, has been proposed. It relies on a new concept of overlap graphs that efficiently overcomes the main difficulty - overlapping occurrences - in probabilities computation. An implementation is available at http://server2.lpm.org.ru/bio/online/sf/. This algorithm provides a significant space improvement over a previous algorithm, `AhoPro` developed with our former associate team MIGEC. Word statistics were used to identify mRNA targets for miRNAs involved in carcinogenesis [13].

Large deviation results have been derived in [41] that take advantage of general combinatorial properties of words. First, an approximation is derived for the double strands counting problem that refers to a counting of a given pattern in a set of sequences that arise from both strands of the genome. Here dependencies between a sequence and its complement plays a fundamental role. Second, sets of small sequences, with non-identical distributions, are addressed. Possible applications are the search of cis-acting elements in regulatory sequences that may be known, for example from ChIP-chip or ChipSeq experiments, as being under a similar regulatory control.

In [21], we developed a new algorithm for generating uniformly at random words of any regular language L. When using floating point arithmetics, its bit-complexity is $O(q \log n)$ in space and $O(qn \log n)$ in time, where n stands for the length of the word, and q stands for the number of states of a finite deterministic automaton of L. We implemented the algorithm and compared its behavior to the state-of-the-art algorithms, on a set of large automata from the VLTS benchmark suite. Both theoretical and experimental results show that our algorithm offers an excellent compromise in terms of space and time requirements, compared to the known best alternatives. In particular, it is the only method that can generate long paths in large automata. Moreover, in [10], in collaboration with the Fortesse group at LRI, we presented several randomised algorithms for generating paths in large models according to a given coverage criterion. This work opens new perspectives for future studies of statistical testing and model checking, mainly to fight the combinatorial explosion problem.

5.3.2. Analysis and design of weighted combinatorial models

Weighted context-free grammars are natural – yet powerful – random models for biological sequence and structures. We furthered our developments on these objects, and applied them to the study of the Boltzmann ensemble of low-energy in RNA.

In collaboration with P. Clote (Boston College), we used such analytic combinatorics to establish that the average geometric distance between the terminal ends of an RNA sequence, once folded, is asymptotically constant [8].

Furthermore, in collaboration with C. Banderier, O. Bodini and H. Tafat (LIPN), we constructively showed that any predefined distribution of pattern could be attained by a (possibly ambiguous) regular expressions. We also designed a dynamic-programming algorithm to automatically build such models, adopting a segmentation approach based on a parsimony principle. This work was presented at the ANALCO'12 conference [30].

Finally, we continued with D. Gardy and J. Du Boisberranger (PRISM, Université de Versailles-St Quentin) a joint study of collisions in weighted random generation. Indeed, while performing a random generation within large collections of weighted objects, the probability of any sample can be exactly and efficiently computed. Therefore, any redundancy in the sampled set is uninformative (contrasting with situations where the probability is also estimated by the sampling procedure). Following previous results presented at GASCOM'10 (Montreal), we presented at the AOFA'12 (Montreal, Canada) conference [33], a new close formula for the waiting-time of the coupon collector problem, i.e. the average number of words that one must draw to obtain the full collection. The framework defined here has direct applications in the context of RNA : approaches based on sampling are preferred to deterministic optimizations, and algorithmic efficiency of the methods can be critically affected by the redundancy of sampled sets. .

5.3.3. Scientific Workflows

Several Scientific workflow systems have been designed to support users in the tasks of designing, managing, monitoring, and executing in-silico experiments. Such systems are now equipped of provenance modules able to collect data produced and consumed during workflow runs to enhance reproducibility. In this context, we have worked in two directions. First, we have worked on the problem of reuse between scientific workflows. In particular, we have identified the presence of common or similar (sub-)workflows and workflow elements, and have deeply studied, for the first time in the literature, the problem of cross-author reuse [38].

Second, we have worked on studying the structure of scientific workflows. More precisely, we have focused on the series-parallel graph structures. Designing sub-workflows, querying or monitoring workflows leads to perform graph sub-isomorphism. This problem is NP-complete when general DAGs are considered but can be solved in polynomial time when graphs restricted to SP graphs are considered. We have designed and implemented the SPFlow algorithm that rewrites any workflow into an SP workflow while ensuring that the provenance of the rewritten workflow is the same as the original [32], [39].

We are currently working on identifying the reasons why some scientific workflows have a non SP structure. Our long-term goal is to design a *distilling procedure* for scientific workflows offering users the ability of naturally designing workflows having a structure close to SP structures. This work is done in close collaboration with the University of Manchester [31].

5.4. Systems Biology

5.4.1. Reasoning on knowledge to build signaling networks:

We have introduced a logic-based method to infer molecular networks and show how it allows inferring signalling networks from the design of a knowledge base. Provenance of inferred data has been carefully collected, allowing quality evaluation. Our method (i) takes into account various kinds of biological experiments and their origin; (ii) mimics the scientist's reasoning within a first-order logic setting; (iii) specifies precisely the kind of interaction between the molecules; (iv) provides the user with the provenance of each interaction; (v) automatically builds and draws the inferred network [29].

5.4.2. Metabolic pathways

The topological analyse of metabolic networks is a first step to understand their behaviours and is described in term of fluxes analyses. We work on the elaboration of a stoichiometric model of *Bacillus subtilis* where its fluxes analyse predicted transcriptional regulation to be more important for the dynamics induced by glucose than by malate [7].

In metabolic pathway analyses, the metabolic networks are described in term of biochemical reactions and metabolites. The integration of structural data is required for a comprehensive understanding of the metabolic networks. We represent the metabolic networks with the functional connectivity between the protein functional domains to make more relevant analyses. We used BioΨ, a formal multi-level description based on elementary actions, to assign functions on structural domains and the elementary flux modes theory to check if the already known pathways remain presents and to identify new ones.

A new version of the software has `Mpas` (Metabolic Pathway Analyser Software) been developed during a Master2'internship by Gh. Fievet. Meanwhile we have also introduced in the landscape of the cell its membranes and the numerous pumps that facilitate ions transfers, hence taking into account the pH of the cytoplasm, a parameter that fits the cell mytosis cycle and which proves to separate the cancerous/normal status of cells [22]. We now aim at study larger and more elaborate metabolic systems, including the Krebs cycle and the mitochondria influence, thus enhancing the scalability of our method [17].

5.4.3. Bacterial phenotypic adaptation

We attempt to re-interpret a major event, the initiation of chromosome replication in *Escherichia coli*, in the light of scales of equilibria. This entails thinking in terms of hyperstructures as responsible for intensity sensing and quantity sensing and how this sensing might help explain the role of the DnaA protein in initiation of replication. We outline experiments and an automaton approach to the cell cycle that should test and refine the scales concept [19].

Another possible direction to study the mechanisms used by cells to integrate and respond to their environment is to search for a link between two large hyperstructures: the cytoskeleton and the general metabolic activity of the cell. There is extensive evidence for the interaction of metabolic enzymes with the eukaryotic cytoskeleton. We state the hypothesis that the cytoskeleton senses and integrates the general metabolic activity of the cell. The physical and chemical effects arising from metabolic sensing by the cytoskeleton would have major consequences on cell shape, dynamics and cell cycle progression. The hypothesis provides a framework that helps the significance of the enzyme-decorated cytoskeleton be determined [18].

In order to test these hypotheses, we have added many features to the HSIM simulation software. The main addition being a way to get both the power of expression of the "entity-centred" paradigm and the computational efficiency of global methods, such that Gillespie-like stochastic simulation algorithm (SSA). To achieve this, we have implemented two new algorithms. The first one concerns the possibility to take into account the interactions between two classes of molecules: the one we want to follow the spatial location over time (entities) and the one for which only the evolution of the number of copies over time is relevant.

The second algorithm is an enhancement of the tau-leap variant of the exact Gillespie SSA; This allows to take into account the interactions between globally treated molecules. The HSIM-SSA algorithm performs an adaptive processing of the number of reactions which may have been triggered during the time step. At each time step, the fast reactions are averaged while the slow reactions are fully stochastically treated. This allows HSIM-SSA to be more than 10 times faster than the other tau-leap SSA implementations [28].

5.4.4. Use of bacteria for biotechnology

Another center of interest has been to find a way to use bacteria as a mean to help us to engineer new biomolecules with specific characteristics. It is sometimes speculated that the equivalent of the polymerase chain reaction might be developed for identification of peptides, proteins or other molecules. Natural amplification systems do exist as in the case of certain autoinducer systems in bacteria. We have been outlined a possible, generic method, *the mimic chain reaction*, for obtaining peptides with 3-D structures that mimic the 3-D structure of their targets. These targets would include a variety of molecules, including proteins. There are therefore two categories of applications: the ability via amplification firstly to detect a known protein or other target at an extremely low concentration and secondly to obtain a set of peptides that mimic the structure of an unknown target and that can be used to obtain a *photofit* [20].

ASCLEPIOS Project-Team

5. New Results

5.1. Medical Image Analysis

5.1.1. Brain tumor cell density estimation from multi-modal MR images based on a synthetic tumor growth model

Participants: Ezequiel Geremia [Correspondant, Inria], Nicholas Ayache [Inria], Antonio Criminisi [MSRC], Bjoern Menze [Inria,ETHZ], Marcel Prastawa [University of Utah].

Published in the proceedings of the MCV Workshop at MICCAI 2012 [35]

biophysiological tumor growth simulator, multi-variate regression random forests, gliomas, MRI

- A generative-discriminative framework is presented to learn model-based estimations of the tumor cell density
- The ground truth for tumor cell density is very hard to obtain
- A biophysiological tumor growth simulator is used to generate the ground truth tumor cell densities and associated MRIs
- A multi-variate regression random forests is trained to estimate the voxel-wise distribution of tumor cell density from input MR images
- The training data contains 500 synthetic cases and their associated ground truth generated by the brain tumor simulator
- The method was tested on 200 synthetic cases with excellent results
- The method also provided very promising results for estimating the tumor cell density on 16 clinical cases showing low grade gliomas from the DKFZ (German Cancer Research Center)

5.1.2. Automatic indexation of cardiac MR images

Participants: Jan Margeta [Correspondant], Nicholas Ayache, Antonio Criminisi [MSRC].

This work has been partly supported by Microsoft Research through its PhD Scholarship Programme and the European Research Council through the ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Machine learning, Cardiac MR, MR preprocessing

- A generic random forest framework has been implemented and its recent modifications have been applied to a fully automatic and a semisupervised image segmentation methods, and manifold learning in cardiac MRI.
- We have performed image based cardiac function quantification from preprocessed cardiac cine MRI sequences.
- An image intensity standardization in magnetic resonance images method has been proposed.

5.1.3. Multimodal brain tumor segmentation

Participant: Bjoern Menze [Correspondant].

MICCAI 2012

- Further developed the generative brain tumor segmentation model
- Developed a generative-discriminative model for multimodal brain tumor segmentation
- Developed a new regularization approach for longitudinal tumor segmentation (with Guillaume Charpiat & Yuliya Tarabalka, Inria Sophia-Antipolis)
- Initiated and co-organized an international benchmark on multimodal brain tumor segmentation as a challenge workshop during MICCAI 2012 in Nice (http://www2.imm.dtu.dk/projects/BRATS2012)

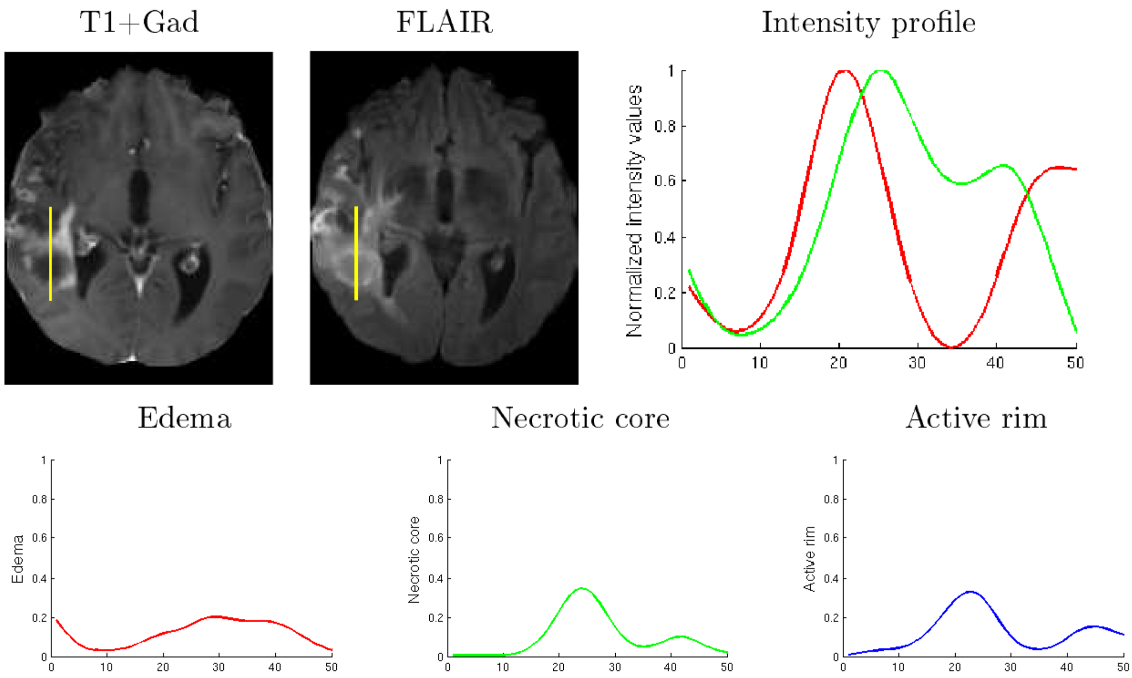


Figure 1. ***Prediction of the cell densities along a section of the tumor.*** *Top, from left to right: T1+Gadolinium, FLAIR image, the intensity profile along the section (yellow). Bottom, from left to right: prediction of the cell density for the edema, necrotic core and active rim, respectively.*

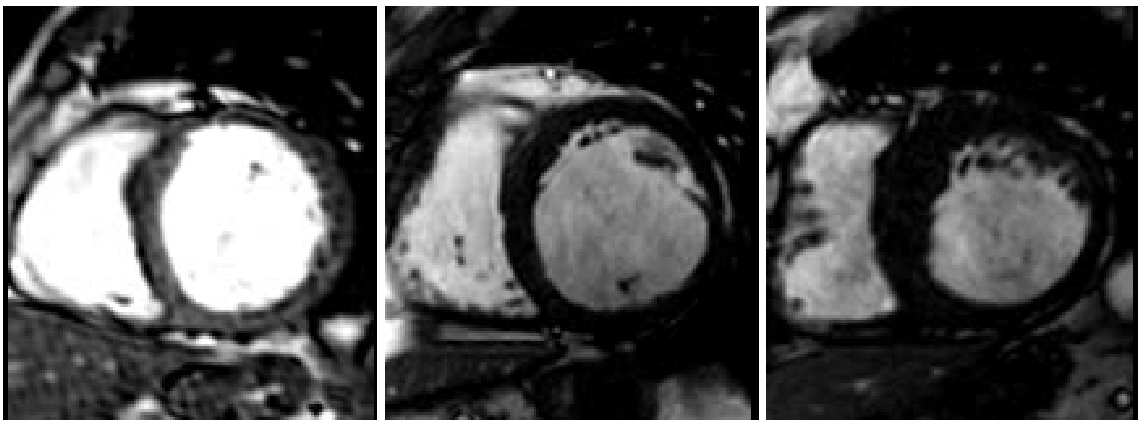


Figure 2. Three different patients with increasing (from left to right) functional deficiency of the heart.

5.1.4. Statistical Analysis of Diffusion Tensor Images of Brain

Participants: Vikash Gupta [Correspondant], Xavier Pennec, Nicholas Ayache.

Diffusion Tensor Imaging of Brain, Tractography, Super-resolution, Statistical analysis

Diffusion tensor imaging (DTI) is gaining interest as a clinical tool for studying a number of brain diseases pertaining to white matter tracts and also as an aid in neuro-surgical planning. Unfortunately, in a clinical environment, the diffusion imaging is hampered by the long acquisiton times, low signal to noise ratio and a prominent partial volume effect due to thick slices. The present work aims at robustifying the analysis of clinical images by developing a super-resolution algorithm for DTI and quantifying its improvements with respect to the existing tensor estimation methods. Part of the work was presented at the 1st International Symposium on Deep Brain Connectomics [68].

5.1.5. 3D/2D coronary arteries registration

Participants: Thomas Benseghir [Correspondant], Grégoire Malandain, Régis Vaillant [GE-Healthcare], Nicholas Ayache.

This work is done in collaboration with GE-Healthcare (Buc).

3D/2D registration ; coronary arteries ; Chronic Total Occlusion ; X-ray fluoroscopy / CT image fusion

The context of this work is to provide the cardiologist with an advanced guidance application, where a pre-operative 3D CT segmented image will be superimposed on the per-operative 2D live fluoroscopy. Since the relative positions of the 3D image and the 2D projective images are unknown, we are currently investigating robust pose estimation methods before using an upcoming registration algorithm.

5.2. Biological Image Analysis

5.2.1. Pre-clinical molecular imaging: breath-hold reconstruction in micro-SPECT and segmentation of IHC stomach slices

Participants: Marine Breuilly [Correspondant], Grégoire Malandain, Nicholas Ayache, Jacques Darcourt [CAL], Philippe Franken [CAL], Thierry Pourcher [CEA].

This work is jointly conducted with the Transporter in Imagery and Oncologic Radiotherapy team (TIRO, CEA-CAL-UNSA) located in Nice.

SPECT/CT, small animal, respiratory motion, respiratory gating, 4D images, stomach, segmentation, immuno-histochemistry

Using the coupled CT and SPECT device, both the anatomy (with the CT) and physiology information targeted by a dedicated radio-pharmaceutical tracer (here the tumors, with the SPECT) can be imaged. However, tumor quantification is impaired by the respiratory motion that induces an artifical enlargement of the moving structures. Thus, the characterization of respiratory motion in dynamic images was studied.

- An ad hoc method for motion detection in dynamic image was developped and tested on two different modalities (4D-SPECT and 4D-CT).
- Image-based motion detection results were compared to the pressure signal and to lung volume variation. A temporal shift between the peak of motion in images and the ones in the pressure signal was observed (see Figure 3).
- The temporal shift suggested to carefully select data from the non moving phase for a motionless 3D-SPECT image reconstruction. This step was incorporated in a breath-hold like reconstruction method [65], [67], [66].

5.3. Computational Anatomy

5.3.1. Statistical Analysis of Transformations on Lie groups and longitudinal studies

Participants: Xavier Pennec [Correspondant], Marco Lorenzi, Nicolas Duchateau [Hospital Clinic, Univ. Barcelona].

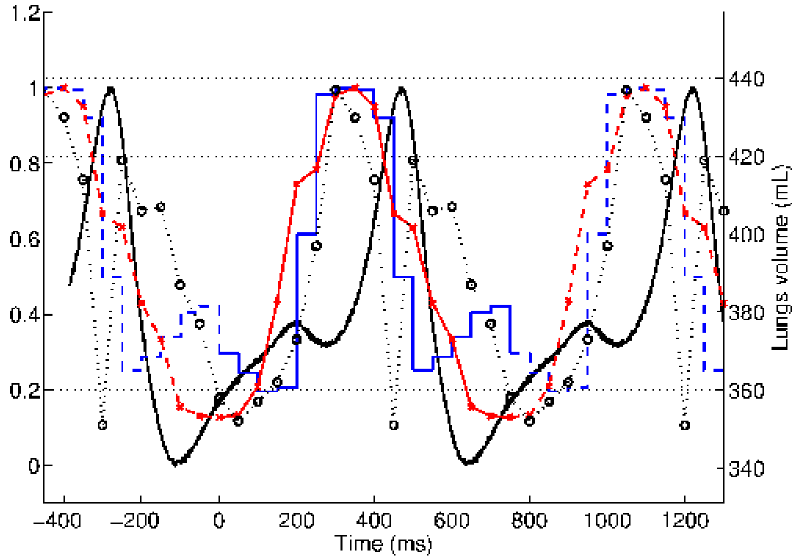


Figure 3. ***Comparison between pressure signal and image-based measures computed on 4D-SPECT and 4D-CT.*** *Normalized average pressure signal (continuous black curve); normalised histogram* $h_{4D-SPECT}(n)$ *computed on the 4D-SPECT image (blue staircase); normalised histogram* $h_{4D-CT}(n)$ *of motion phase computed on the 4D-CT image (red curve with cross-shaped markers); and lung volume (mL) measured on the 4D-CT image (black dotted curve with circle-shaped markers).*

Lie groups, transformations, mean value, non-linear registration

In order to perform statistics on transformations for computational anatomy purposes, we investigate alternative theoretical structures to the right- (or left-) invariant Riemannian setting usually used.

- In order to define a notion of a mean which is consistent with Lie group operations we propose in [57] to replace the Riemannian metric by an affine connection structure on the group. We show that the canonical Cartan connections of a connected Lie group provides group geodesics including one-parameter subgroups which are completely consistent with the composition and inversion. To extend statistical operations to such a non-metric structure, we propose an implicit definition as an exponential barycenter (there is no Fréchet mean like in Riemannian Manifolds) and a linearly convergent iterative fixed point algorithm to reach it. This results into naturally bi-invariant means which are unique when the dispersion of the data is small enough. In some cases including rotations and rigid-body transformations, there is even a global existence and uniqueness theorem which is similar to the Riemannian case.
- In [14], we investigate the canonical Cartan connections and their associated parallel transport for diffeomorphisms, which justifies the use of one-parameter subgroups (the flow of stationary velocity fields or SVF) for diffeomorphic image registration. In particular, we derive closed-forms for different parallel transports and we compare SVF and LDDMM approaches with experiments on longitudinal and inter-subject registration.
- In [33], we analyses with practical expriments what kind of parallel transport is needed to reorient the deformation characteristics along the time sequences of the cardiac motion. Contrarily to the case of the brain, inter-subject transformations to normalize the heart between different subjects are of the same order than deformations along the sequence.

5.3.2. Statistical Analysis of Longitudinal Transformations in the LDDMM framework

Participants: Stanley Durrleman [Correspondant], Xavier Pennec, Alain Trouvé [CMLA, ENS Cachan], Nicholas Ayache, José Braga [UMR 5288 CNRS-Université Toulouse Paul Sabatier].

Lie groups, transformations, mean value, non-linear registration

The work initiated the previous years with the PhD of S. Durrleman on the spatio-temporal modeling of shapes was applied with J. Braga to quantify ontogenetic differences between bonobo (Pan paniscus) and chimpanzee (Pan troglodytes) endocrania, using dental development as a timeline. We perform a temporal surface regression that estimates typical endocranial ontogenetic trajectories separately for bonobos and chimpanzees which highlights non-linear patterns of endocranial ontogenetic change and significant differences between species at local anatomical levels rather than considering the endocranium as a uniform entity. The decomposition of the spatio-temporal inter-species difference into a morphological deformation (accounting for size and shape differences independently of age) and a time warp (accounting for changes in the dynamics of development) indicates that juvenile bonobos develop much slower than juvenile chimpanzees, suggesting that inter-specific ontogenetic shifts do not only concern endocranial volume increase, but also the rate of shape changes over time. Our method provides, for the first time, a quantitative estimation of inter-specific ontogenetic shifts that appear to differentiate non-linearly. This work was pusblished in the journal of human evolution [10].

5.3.3. The Kernel Bundle Framework for Diffeomorphic Image Registration

Participants: Xavier Pennec [Correspondant], Stefan Sommer [Computer Science Dpt, University of Copenhagen, DK], François Lauze [Computer Science Dpt, University of Copenhagen, DK], Mads Nielsen [Computer Science Dpt, University of Copenhagen, DK].

This work in collaboration with the Computer Science Department of the University of Copenhagen (DK) was initiated during the 6 month visit of S. Sommer at Asclepios in 2010-2011 and was continued remotely since then.

non-rigid registration algorithm, statistics, deformations, shapes, locally affine deformations, sparsity

In order to detect small-scale deformations during longitudinal registration while allowing large-scale deformation needed for inter-subject normalization, we wish to model deformation at multiple scales and represent the deformation at the relevant scales only. We combined in [48], [26] a sparsity prior with the multi-scale Kernel Bundle framework, resulting in an algorithm allowing compact representation of deformation across scales.

In [27], we further extend the framework by introducing higher-order momentum distributions in the LDDMM registration framework. While the zeroth order moments previously used in LDDMM only describe local displacement, the first-order momenta that are proposed here represent a basis that allows local description of affine transformations. Beyond the careful mathematical construction, we show the implications for sparse image registration and we provide examples of how the parametrization enables registration with a very low number of parameters.

5.3.4. Spectral Correspondances in Non-linear Image Registration

Participants: Xavier Pennec [Correspondant], Hervé Lombaert, Nicholas Ayache, Leo Grady [SCR, Princeton, US], Farida Cheriet [Saint-Justine Hospital, Montreal, CA].

This work was performed in collaboration with Saint-Justine Hospital in Montreal (CA) and Siemens Corporate Research in Princeton (US).

non-rigid registration algorithm

The demons algorithm was enhanced to include spectral feature correspondences between the images [37]. This feature proves to drastically enhance the robustness of the registration algorithm, which turns out to have a major impact on the construction of atlases. This work was awarded the best paper award at the Medical Computer Vision Workshop [38] and was protected by a patent filling in the US [62]

5.3.5. Longitudinal Analysis of Brain Atrophy in Alzheimer's Disease

Participants: Marco Lorenzi [Correspondant], Xavier Pennec, Nicholas Ayache, Giovanni B. Frisoni [IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy].

This work is done in collaboration with LENITEM, IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy.

Alzheimer's Disease, non-rigid registration algorithm, longitudinal analysis.

The accurate analysis of the longitudinal structural changes in the brain plays a central role in the study of Alzheimer's disease (AD), for diagnostic purposes and for the assessment of the drugs efficacy in clinical trials. The goal of this project is to provide robust and effective instruments based on non-rigid registration of serial MR images for the modeling and the quantification of the brain atrophy evolution in AD. In 2012, our main scientific developments were the following:

- We developed a framework for the consistent definition of anatomical regions of longitudinal brain atrophy, and for the robust quantification of longitudinal regional percentage volume loss. The framework is based on the analysis of the flux associated to longitudinal deformations (see Figure 4), and was successfully applied to large public dataset of brain images (ADNI - http://adni.loni.ucla.edu/). The work was accepted for oral podium presentation at the MICCAI conference 2012 [40].
- We applied the flux analysis for the quantification of the longitudinal hippocampal and ventricular atrophy in AD. The proposed framework was presented at the NIBAD MICCAI Challenge 2012 [41] , and compared favorably with state-of-art methods in terms of accuracy and stability when applied on the challenge dataset.
- We proposed in [39] a model of the morphological changes in Alzheimer's based on the disentangling of the normal aging component from the pathological atrophy. The model was promoted and presented to the neuroscience community during international scientific conferences [71], [70].

These scientific advances were also included along with the previous ones in the PhD manuscript [1].

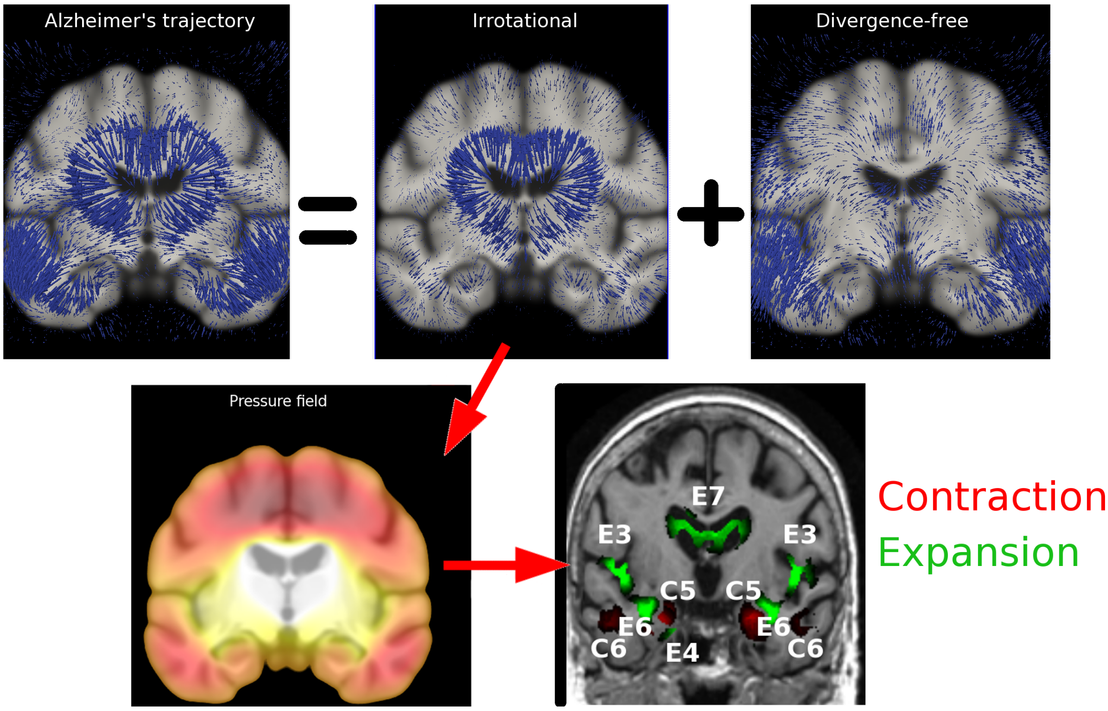


Figure 4. Flux analysis of the longitudinal morphological changes in Alzheimer's disease. The irrotational component associated to the longitudinal deformation encodes the volume loss and gain (top figure). This component is parameterized by a pressure potential (bottom left) which is used to determine the areas of significant flux of the deformation (matter expansion and contraction) in patients affected by Alzheimer's disease (bottom right).

5.3.6. Statistical Modelling of Cardiac Growth, Deformation and Blood Flow from Medical Images

Participants: Kristin McLeod [Correspondant], Adityo Prakosa, Christof Seiler, Maxime Sermesant, Xavier Pennec.

This work was partially funded by the EU project Care4me ITEA2.

Image registration, Demons algorithm, LDDMM, reduced models, CFD, polyaffine, cardiac motion tracking

This work involves developing reduced models of cardiac motion, blood flow and growth.

- Extending the 2011 motion tracking challenge [43], the iLogDemons registration algorithm was applied this year to a data-set of synthetic echocardiography sequences with a training set (provided with ground truth) and testing set to quantitavely compare this algorithms with other cardiac motion tracking algorithms [45].
- A reduced order model of cardiac motion based on a polyaffine log-demons registration was developed to represent the motion along the cardiac cycle with a smaller number of parameters compared to previously proposed methods. The method was applied to a data-set of 10 volunteers and the results were presented at the 2012 STACOM workshop at the MICCAI conference [44].
- The analysis of a statistical model for reduced blood flow simulations in the pulmonary artery proposed in the 2010 STACOM workshop is currently being extended to a journal version with an improved method and a larger data-set.
- The statistical modeling of the right ventricle growth in a population of Tetralogy of Fallot patients was extended to a full bi-ventricular growth model on different data [55]. Results confirm the previous findings which were shown to be useful in providing insights for patient treatment [13] (see Figure 5).

5.3.7. Trees on Geometrical Deformations to Model the Statistical Variability of Organs in Medical Images

Participants: Christof Seiler [Correspondant], Xavier Pennec, Mauricio Reyes [Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland].

This work is performed in the context of the joint PhD of Christof Seiler at the Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland and Asclepios Inria [3].

Parametrization of diffeomorphisms, Shape statistics, Multiscale and hierarchical trees, Log-Euclidean polyaffine transformations, Polyaffine registration, Log-Demons registration, Generative statistical model, Bayesian registration, Mandibles, Femurs

Intersubject anatomical deformations between patients can be found on coarse and fine scales. Each level of granularity has specific regions of interest in clinical applications. The challenge is to connect geometrical deformations to clinical regions across scales.

- We presented this connection by introducing structured diffeomorphic registration [24]. At the core of our method is the parametrization of geometrical deformations with trees of locally affine transformations describing intersubject variability across scales (see Figure 6).
- The methodology of [24] was successfully applied to mandible implant design [31] and in a clinical journal paper on allograft selection [21].
- We statistically modeled the deformation parameters in a population by formulating a generative statistical model [47]. This model allowed us to incorporate deformation statistics as a prior in a Bayesian setting and it enabled us to extend the classical sequential coarse to fine registration to a simultaneous optimization of all scales.
- We explored cell shape statistics to classify stem cells [23].
- We investigated the benefits of considering patient metadata and morphometric measures to enhance bone surface shape prediction [6].

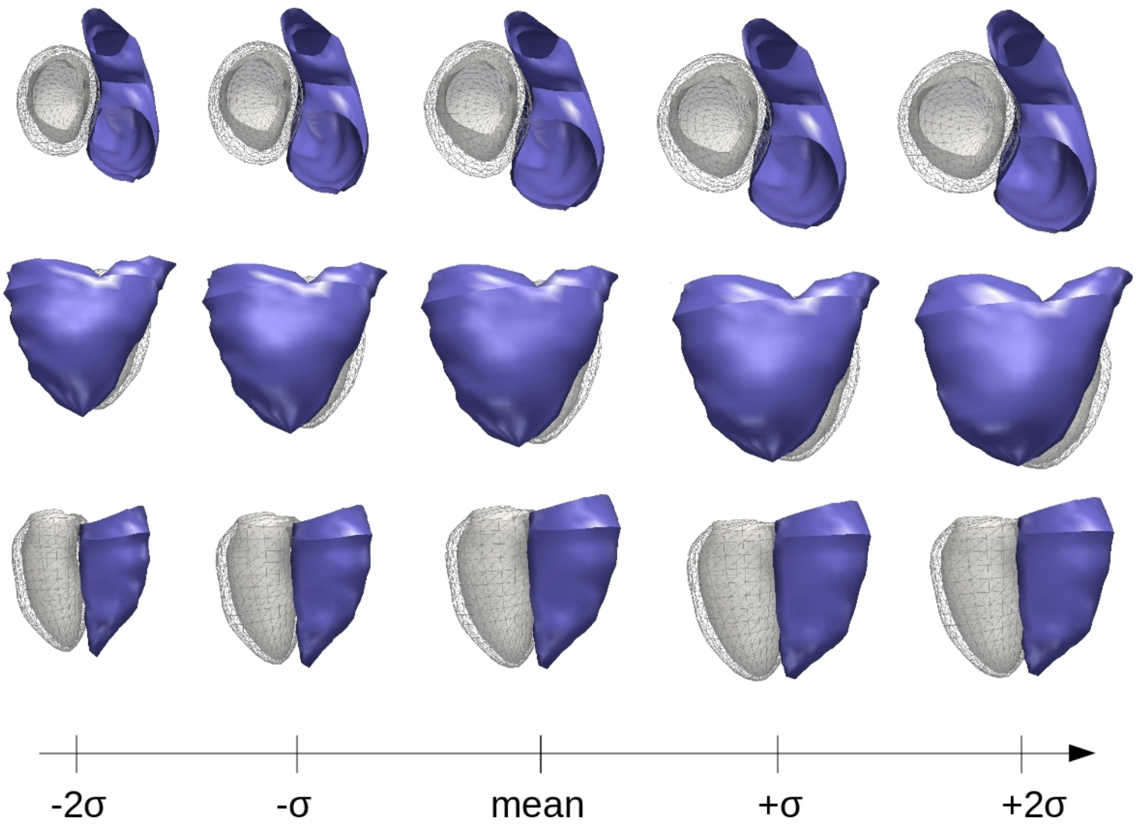


Figure 5. Mean growth model computed from a population of 13 repaired Tetralogy of Fallot patients. Both ventricles grow as body surface area (BSA) increases.

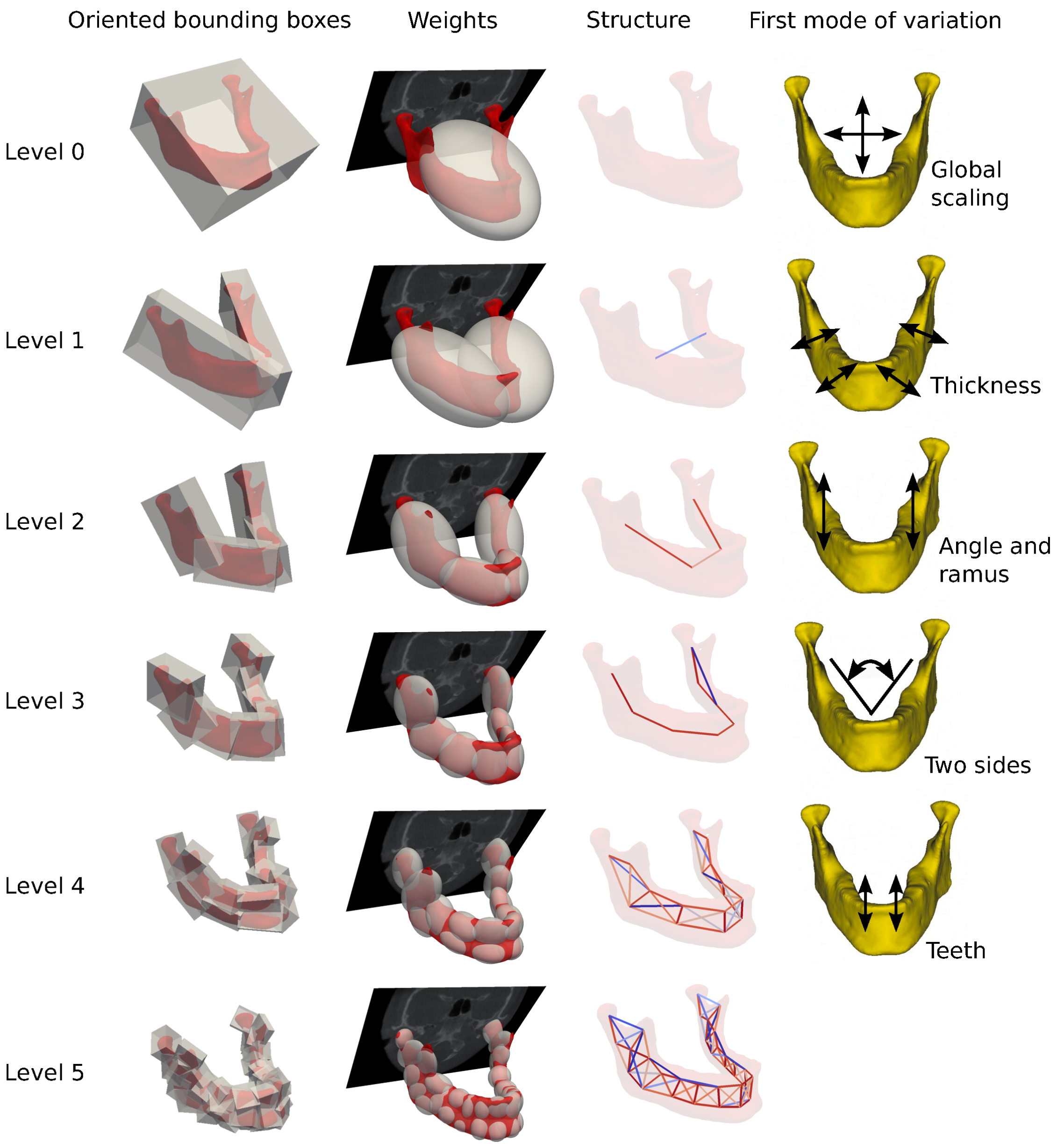


Figure 6. The red mandible is the surface extracted from the template image. Column 1: Oriented bounding boxes computed using the algorithm presented in [24]. Column 2: Ellipsoids representing Gaussian weights derived from the oriented bounding boxes. Column 3: Structure given by the weights with correlations between regions ranging from low (blue=0:4) to high (red=1). Column 4: First PCA mode at each level showing the residual variation.

5.3.8. Evaluation of iLogDemons Algorithm for Cardiac Motion Tracking in Synthetic Ultrasound Sequence

Participants: Adityo Prakosa [Correspondant], Kristin McLeod, Maxime Sermesant, Xavier Pennec.

This work was partially funded by the European Research Council (ERC) through the support of the MedYMA advanced grant 291080 and the European project euHeart.

synthetic echocardiography, iLogDemons, cardiac motion tracking

- The LogDemons and iLogDemons non-linear registration algorithms were evaluated on a dataset of synthetic cardiac ultrasound sequences [32], [45]. With these synthetic sequences, it is possible to quantify the performance of these registration algorithms since the ground truth motion was given. Therefore the LogDemons/ iLogDemons can be evaluated objectively (see Figure 7).

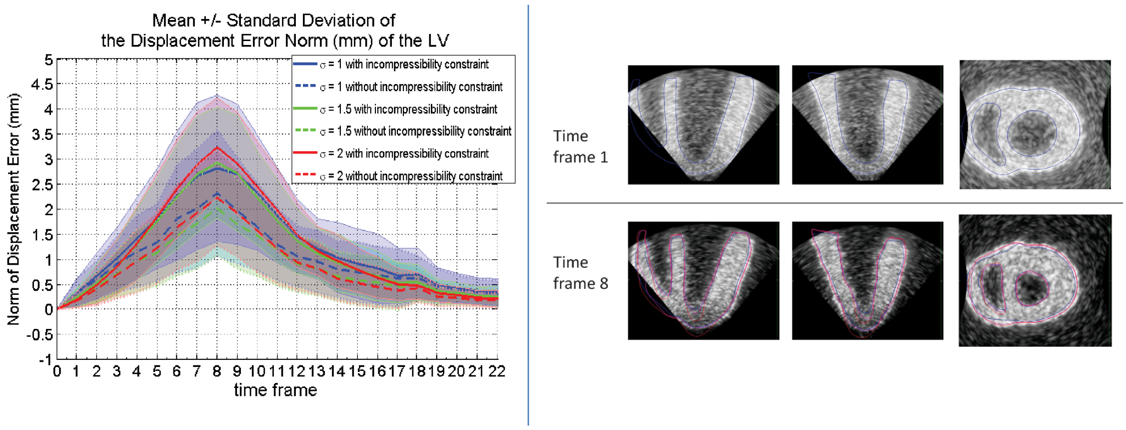


*Figure 7. **Registration Error and Motion Tracking Result** The error quantification for different registration parameter (left) and the myocardium tracking result during the maximum contraction (right): ground truth, LogDemons and iLogDemons estimation are shown in blue, purple and red respectively.*

5.3.9. Simulation of Atrophy in Alzheimer's disease

Participants: Arnaud Le Carvennec [Correspondant], Sebastien Ourselin [UCL], Nick Fox [UCL], Xavier Pennec [Inria], Nicholas Ayache [Inria].

Thesis in collaboration between Asclepios team at Inria and Center for Medical Image Computing (CMIC)-Dementia Research Center (DRC) at University College London (UCL).

Simulation, Alzheimer's disease, registration

- Evaluation of registration algorithm using multi-channel images.
- Simulation of atrophy based on registration.

5.4. Computational Physiology

5.4.1. Tumor Growth Modeling

Participants: Erin Stretton [Correspondant], Emmanuel Mandonnet, Bjoern Menze, Hervé Delingette, Nicholas Ayache.

This work was funded by Care4me, EU program.

DTI, MRI, simulation, clinical, tumor, brain, glioma

We aim at developing image analysis methods and biophysical models in order to guide the planning of therapies (surgical removal and radiotherapy) for brain cancer (glioma) patients. Our work is focused on those objectives :

- Predicting the location of glioma recurrence after a resection surgery [49].
- Determining the best description of tumor cell diffusion tensor in white matter (patient-based, atlas-based or isotropic) which leads to the most accurate results for predicting future tumor growth.
- Comparing tumor growth speeds on 3 patient cases. This is a work in progress and the objective is to reach 30 patients when the work is complete.

5.4.2. Generation of Synthetic but Visually Realistic Time Series of Cardiac Images Combining a Biophysical Model and Clinical Images

Participants: Adityo Prakosa [Correspondant], Maxime Sermesant, Hervé Delingette, Stéphanie Marchesseau, Eric Saloux [CHU Caen], Pascal Allain [Philips Healthcare], Nicolas Villain [Philips Healthcare], Nicholas Ayache.

This work was done in collaboration with Medisys, Philips Healthcare Suresnes, France, and the Cardiology Department of CHU Caen, France. This work was partially supported by the European Research Council through the ERC Advanced Grant MedYMA on Biophysical Modelling and Analysis of Dynamic Medical Images and the European project euHeart.

synthetic 4D cardiac sequences, cardiac electromechanical model, non-rigid registration

- A pipeline to create visually realistic synthetic 4D cardiac sequence using the cardiac motion simulated by an electromechanical model is developed. This pipeline combines the simulated myocardium displacement field with the estimated myocardium displacement field from a registration method. This combined displacement field is then used to warp the original images in order to create the synthetic cardiac sequence.
- In [19],we proposed a new approach based on Stationary Velocity Fields to combine the two motions (see Figure 8). We also proposed a new method that diffuses displacement fields in order to maintain the continuity between the simulation and the real image with minimal texture distortion. Thanks to the detailed interplay between image processing and biophysical modeling, we can fully use a complete sequence in order to generate several new ones. This method also gives better realism compared to traditional methods based on the deformation of an end-diastolic image, since the generated synthetic sequence will also contain the motion of surrounding tissues such as the motion of the mitral valve.
- The new synthetic images are similar to the original ones except for the motion of the heart which is modified to follow the motion provided by a biophysical model. The parameters of the biophysical model can be modified to create variations around this motion. This pipeline has been applied to generate different synthetic sequences from different imaging modalities. It is generic and can be used with a different biophysical model or a different image registration algorithm, and it can be extended to other organs.
- As these synthetic 4D cardiac sequences have kinematic ground truth information, those sequences represent in themselves a valuable resource to benchmark motion tracking methods or to train machine-learning algorithm.

5.4.3. Real-Time Cardiac Electrophysiology Computing for Training Simulator

Participants: Hugo Talbot [Correspondant], Hervé Delingette, Stephane Cotin, Maxime Sermesant, Christian Duriez.

This work was funded by the ADT Sofa and is conducted in collaboration with project teams Shacra and Evasion.

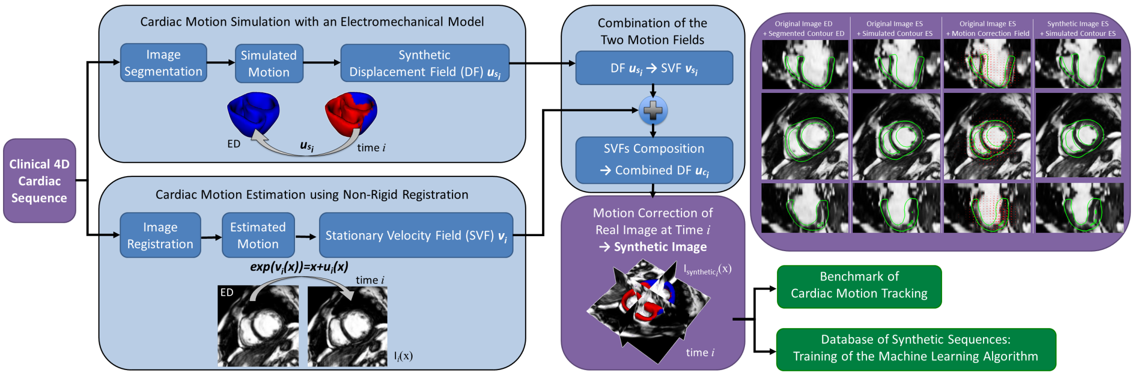


*Figure 8. **Synthetic 4D Cardiac Sequence Generation Pipeline** A clinical 4D sequence is used as an input to create a synthetic 4D sequence in which the myocardium motion follows a prescribed simulated displacement field. The combined simulated and registration motion are used to correct the motion of the real clinical images in order to create the synthetic cardiac sequence.*

Cardiac electrophysiology simulation, real-time, GPU computing, patient-specific study

Cardiac arrhythmia is a very frequent pathology related to an abnormal electrical activity in the myocardium. This work aims at developing a training simulator for interventional radiology and thermo-ablation of these arrhythmias.

- The latest improvements lead on electrophysiology simulation (see Figure 9) using GPU computing allowed us to reach real-time performance[51]. The issue of fast electrophysiology was a major bottleneck in the development of our simulator.

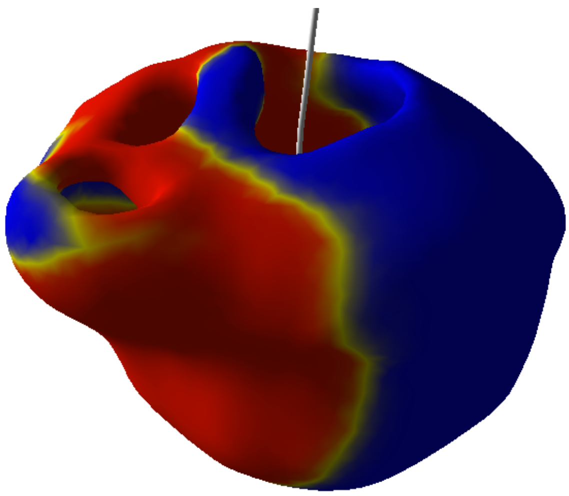


Figure 9. Cardiac electrophysiology computed on a patient-specific geometry

Coupling between the cardiac electrophysiology model with cardiac mechanical models has been achieved, thus leading to an interactive framework. Moreover, the electrophysiology simulation has been also coupled with a navigation simulation.

- In the context of his work on cardiac electrophysiology, we initiated two different collaborations. Joint work has been performed with the team CARMEN from Inria Bordeaux on bidomain modeling for cardiac electrophysiology. This exchange targeted the implementation in SOFA of these models. Secondly, a collaboration with the MACS team in Saclay has been initiated to personalize the cardiac electrophysiology model based on the Verdandi library.

5.4.4. Personalized model of the heart for cardiac therapy planning

Participants: Stéphanie Marchesseau [Correspondant], Hervé Delingette, Nicholas Ayache, Maxime Sermesant.

An award has been won for this work at the MICCAI 2012 Conference. It was partially funded by the European Community's euHeart project under grant agreement 224495 and by the ERC advanced Grant MedYMA 291080.

Cardiac simulation, sensitivity analysis, calibration algorithm, specificity study

- We implemented the full Bestel-Clement-Sorine electromechanical model of the heart in SOFA [54], [51].
- We ran a complete sensitivity analysis to check its behaviour for healthy and pathological cases [15].
- A new calibration algorithm was proposed [15] in order to initialize global mechanical parameters from the volume and pressure curves, before further personalization (see Figure 10).
- The application of this new method on 6 healthy and 2 pathological cases allowed to draw preliminary conclusions on specific parameters to a given pathology [42], [16].
- The model has also been used to create synthetic images in [19] and for the data of the STACOM 2012 challenge [32].

5.4.5. Image-based glioma modeling for radiotherapy planning

Participants: Bjoern Menze [Correspondant], Ender Konukoglu [MSR Cambridge], Jan Unkelbach [Harvard MGH].

- Implemented the generative tumor segmentation model together with the E. Konukoglu's tumor infiltration model for evaluation at the MGH Department of Radiation Oncology.
- Integrated tumor infiltration model with radiation therapy model.

5.4.6. Cardiac Arrhythmia Radio-frequency Ablation Planning

Participants: Rocio Cabrera Lozoya [Correspondant], Maxime Sermesant, Hervé Delingette, Nicholas Ayache.

This work is performed in the context of the the PhD of Rocio Cabrera Lozoya in collaboration with the IHU LIRYC Bordeaux and is funded by ERC MedYMA.

- Biophysical model development for the prediction of radio frequency ablation sites for ventricular tachycardias.
- Target site map generation for ablation therapy guidance
- Structural and functional characterization of target sites using 3D imaging and EP measurements through machine learning algorithms (see Figure 11).
- Prediction validation with acquired clinical data

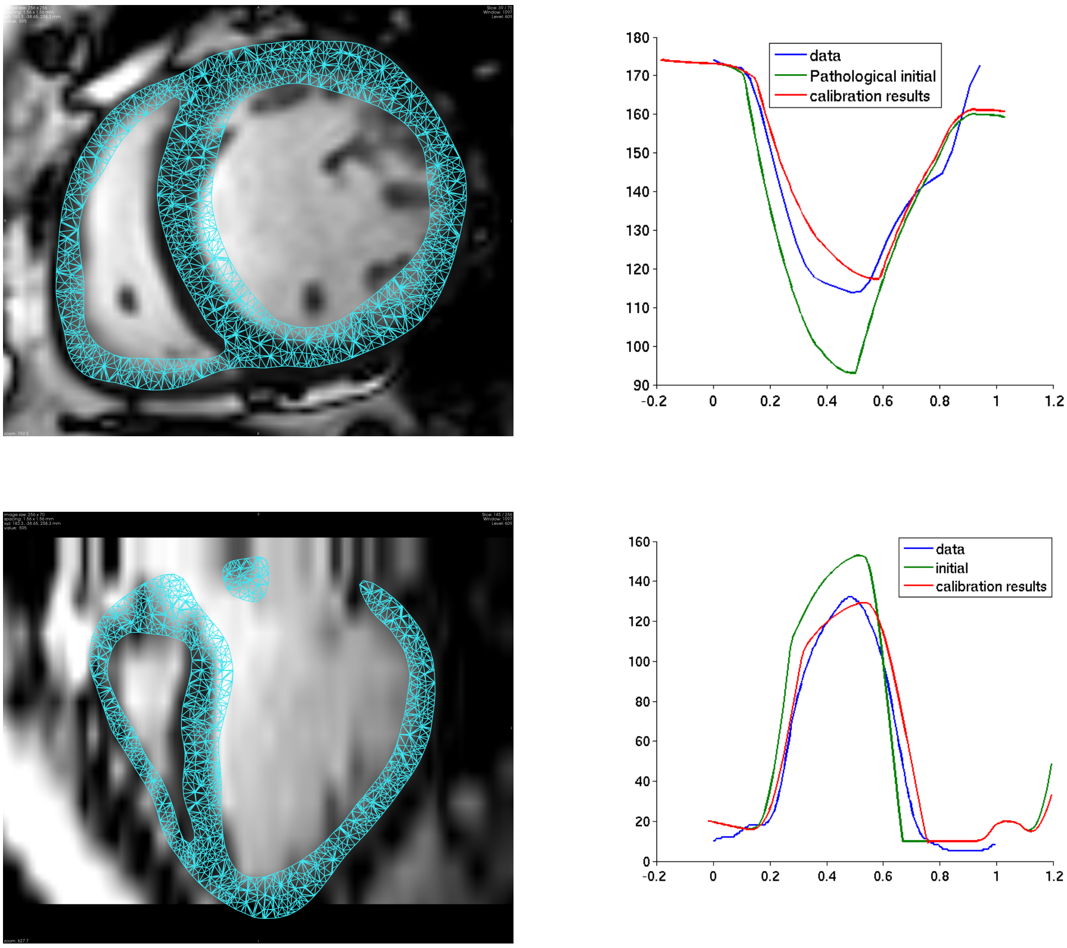


Figure 10. Results of the calibration algorithm fro one pathological case. (Left) Simulated mesh after calibration compared to the images (at end-diastole). (Right) Resulting volume and pressure curves.

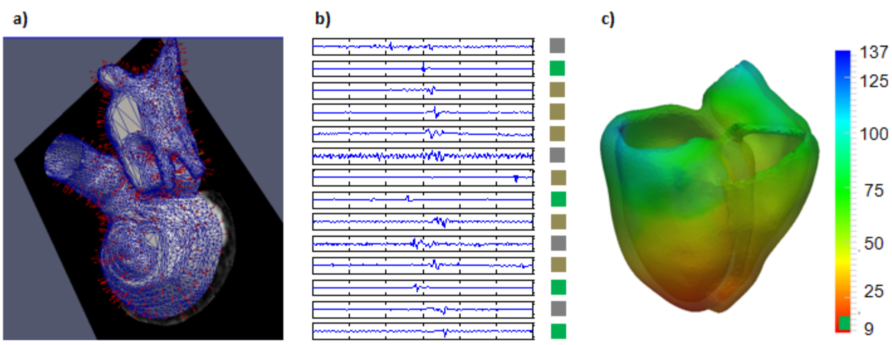


Figure 11. a) MRI Feature Extraction b) EP Signal Feature Extraction c) EP Model Personalization

5.4.7. Computational modeling of radiofrequency ablation for the planning and guidance of abdominal tumor treatment

Participants: Chloe Audigier [Correspondant], Herve Delingette, Tommaso Mansi [Siemens Corporate Research], Nicholas Ayache.

This PhD is carried out between Asclepios research group, Inria Sophia Antipolis, France and the Image Analytics and Informatics global field, Siemens Corporate Research, Princeton, USA.

Therapy planning, radio-frequecy ablation, Liver

The objective of this work is to develop a computational framework for patient-specific planning of radiofrequency ablation:

- A patient-specific detailed anatomical model of the liver is estimated from standard CT image and meshed to generate a tetrahedral volume mesh.
- A porous media model is used to compute the patient-specific blood flow in the hepatic circulatory system.
- Bio-heat equations have been implement in SOFA to model the heat propagation in biological tissues.
- A cell death model is included to account for the cellular necrosis.

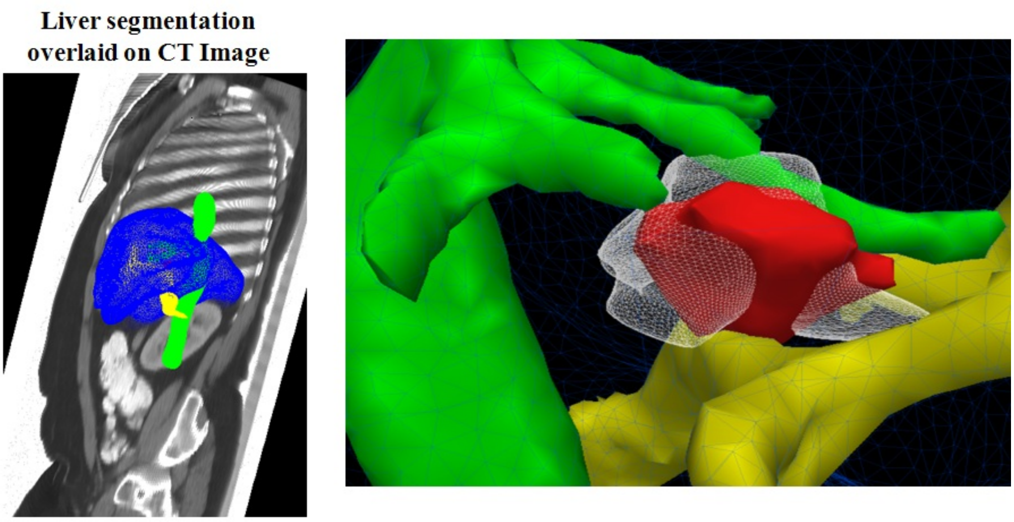


Figure 12. (Left) Anatomical model of the liver estimated from standard clinical CT image. (Right) The predicted necrosis computed with our model compares qualitatively well with the necrosis region observed on a post-operative MRI scan.

5.4.8. Tumor Growth Simulation for the creation of a database of virtual patients

Participants: Nicolas Cordier [Correspondant], Nicholas Ayache, Hervé Delingette, Bjoern Menze, Ezequiel Geremia.

This work was funded by the European Research Council through the ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Brain MRI, Tumor simulation.

- Synthesizing multi-channel MR images with healthy and glial tumors.
- Creating a database of synthetic images for training and validating of brain tumor segmentation algorithms.

5.4.9. Learning approach for the Mechanical personalization of cardiac models

Participants: Loic Le Folgoc [Correspondant], Hervé Delingette, Antonio Criminisi, Nicholas Ayache.

This work was partly funded by Microsoft Research through its PhD Scholarship Programme and by the ERC Advanced Grant MedYMA.

Inverse problem, machine learning, patient-specific, current, kinematics

- A machine-learning framework for the mechanical personalization of the Bestel-Clement-Sorine model of the heart from patient-specific kinematics
- The computational burden is moved to an offline stage, where the inter-subject variability in motion is captured via the statistical analysis of training samples
- Towards a probabilistic framework for the personalization and therapy planning problems, to better account for significant and diverse uncertainty sources
- Published at the MICCAI 2012 Workshop on Medical Computer Vision[36]

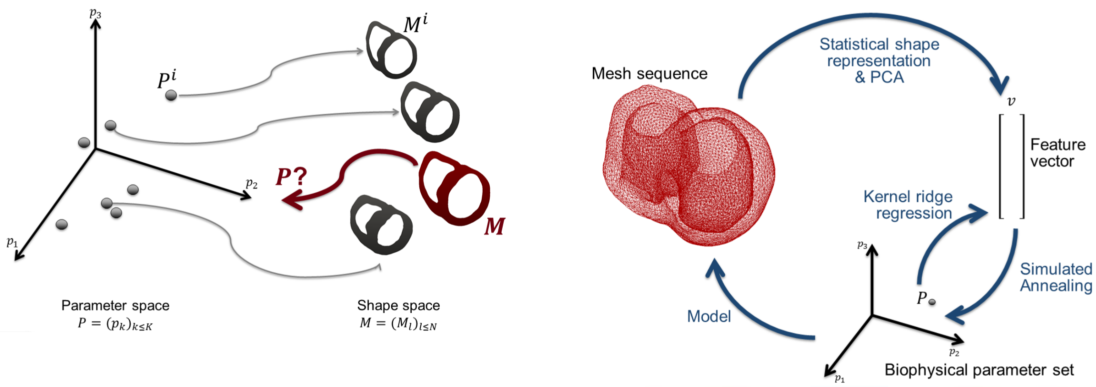


Figure 13.

5.4.10. Brain Tumor Growth Modeling

Participants: Matthieu Lê [Correspondant], Nicholas Ayache, Hervé Delingette.

Gliomas simulations, reaction-diffusion, brain tumors

- In collaboration with the MC2 research team in Bordeaux, we developed a tumor growth model based on different types of cell : necrotic, proliferative and quiescent cells (see Figure 14). It is also based on the underlying vascularization of the brain.
- We studied the impact of the vascularization angiogenesis factor and degradation factor.

5.4.11. Modeling of atrophy of the brain in Alzheimer’s Disease

Participants: Bishesh Khanal [Correspondant], Xavier Pennec, Nicholas Ayache.

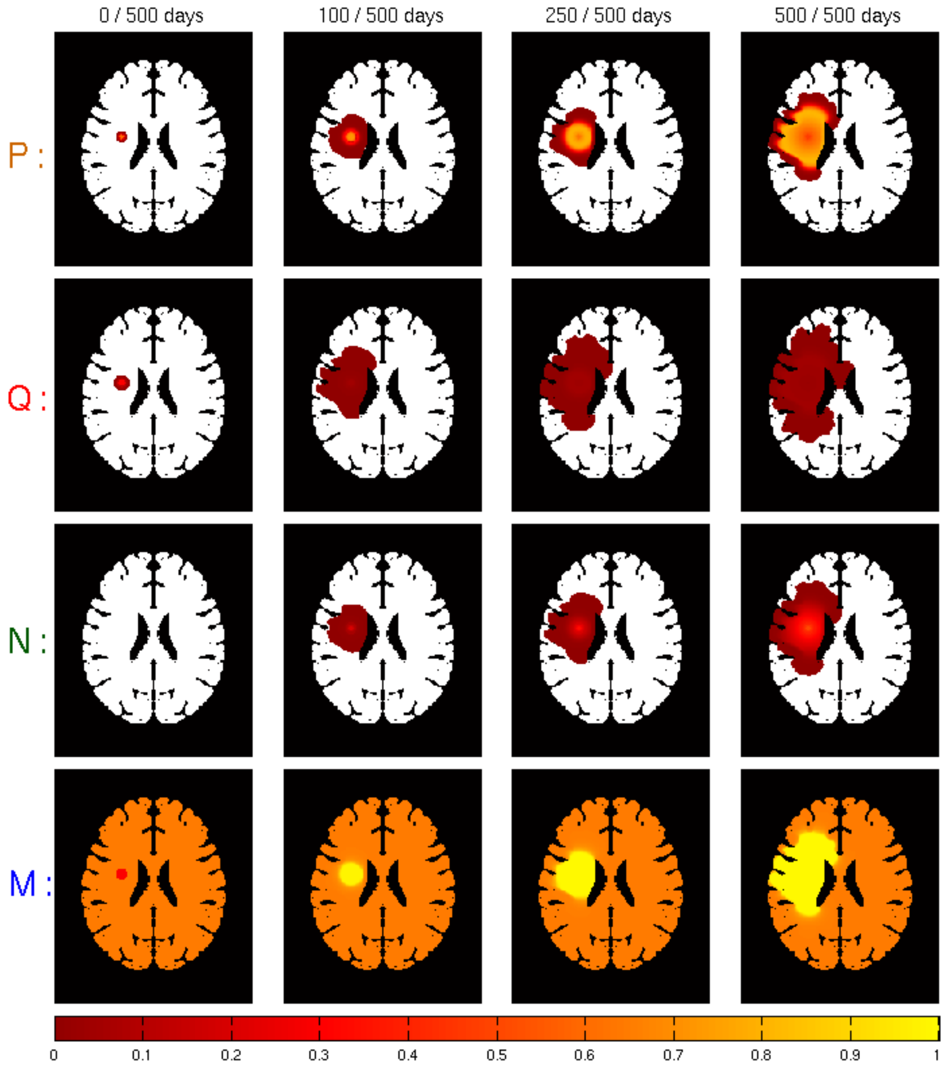


Figure 14. Results of a glioma simulation at day 0, 50, 250 and 500. The proliferative cells are on the first row, the quiescent cells on the second row, the necrotic cells on the third one and the vascularization is on the fourth row.

Alzheimer's Disease (AD), modeling atrophy, biomechanical model

- The idea is to have a model which produces deformation of the brain when a known distribution of local volume change (atrophy) is prescribed to the model. The study is to understand how brain deformation evolve in time with respect to temporal and spatial variation of atrophy.
- During the masters internship period a simple model was tested in 2D square and 3D cube where high atrophy regions acted as sinks for the displacement field [69] (see Figure 15).

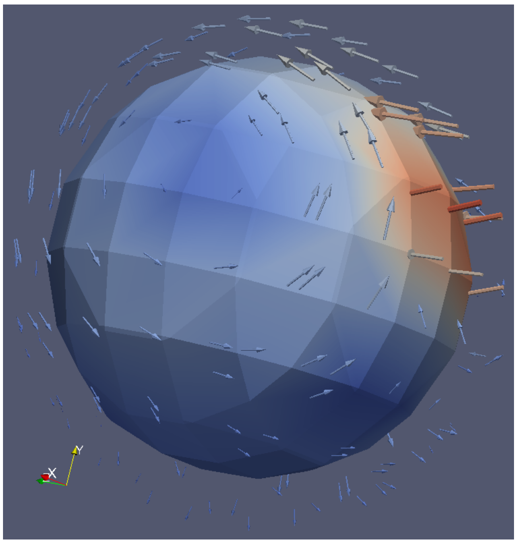


Figure 15. Displacement field in 3D when high atrophy is prescribed in the center of a cube.

ATHENA Project-Team

6. New Results

6.1. Computational Diffusion MRI

6.1.1. Improving dMRI Signal and Acquisitions

6.1.1.1. Diffusion MRI Signal Reconstruction with Continuity Constraint and Optimal Regularization

Participants: Emmanuel Caruyer, Rachid Deriche.

In diffusion MRI, the reconstruction of the full Ensemble Average Propagator (EAP) provides new insights in the diffusion process and the underlying microstructure. The reconstruction of the signal in the whole Q-space is still extremely challenging however. It requires very long acquisition protocols, and robust reconstruction to cope with the very low SNR at large b-values. Several reconstruction methods were proposed recently, among which the Spherical Polar Fourier (SPF) expansion, a promising basis for signal reconstruction. Yet the reconstruction in SPF is still subject to noise and discontinuity of the reconstruction. In this work, we present a method for the reconstruction of the diffusion attenuation in the whole Q-space, with a special focus on continuity and optimal regularization. We derive a modified Spherical Polar Fourier (mSPF) basis, orthonormal and compatible with SPF, for the reconstruction of a signal with continuity constraint. We also derive the expression of a Laplace regularization operator in the basis, together with a method based on generalized cross validation for the optimal choice of the parameter. Our method results in a noticeable dimension reduction as compared with SPF. Tested on synthetic and real data, the reconstruction with this method is more robust to noise and better preserves fiber directions and crossings.

This work has been published in [13]

6.1.1.2. A Computational Framework for Experimental Design in Diffusion MRI

Participants: Emmanuel Caruyer, Rachid Deriche.

In this work, we develop a computational framework for optimal design of experiment in parametric signal reconstruction. We apply this to the optimal design of one dimensional Q-space, Q-ball imaging and multiple Q-shell experimental design. We present how to construct sampling scheme leading to minimal condition number, and compare to state-of-the-art sampling methods. We show in particular a better noise performance of these scheme through Monte-Carlo simulations for the reconstruction of synthetic signal. This demonstrates the impact of this computational framework on acquisition in diffusion MRI.

This work has been published in [16]

6.1.1.3. Parametric Dictionary Learning in Diffusion MRI

Participants: Sylvain Merlet, Emmanuel Caruyer, Aurobrata Ghosh, Rachid Deriche.

This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.

In this work, we propose an approach to exploit the ability of compressive sensing to recover diffusion MRI signal and its characteristics from a limited number of samples. Our approach is threefold. First, we learn and design a parametric dictionary from a set of training diffusion data. This provides a highly sparse representation of the diffusion signal. The use of a parametric method presents several advantages: we design a continuous representation of the signal, from which we can analytically recover some features such as the ODF; besides, the dictionary we train is acquisition-independent. Next, we use this sparse representation to reconstruct the signal of interest, using cross-validation to assess the optimal regularization parameter for each signal reconstruction. The use of cross-validation is critical in the L1 minimization problem, as the choice of the parameter is sensitive to the noise level, the number of samples, and the data sparsity. Third, we use a polynomial approach to accurately extract ODF maxima. Finally, we motivate and describe the choice of experimental parameters for the HARDI contest.

This work has been published in [26].

6.1.1.4. Diffusion and Multiple Orientations from 1.5 MR Systems with Limited Gradient Tables

Participants: Sylvain Merlet, Rachid Deriche, Kevin Whittingstall [Radiology department, Université de Sherbrooke, Québec, Canada], Maxime Descoteaux [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada].

This work has been performed within the framework of the Brain Connectivities Associate Team.

Diffusion MRI (dMRI) enables the quantification of water diffusion, influenced by the structure of biological tissues, from the acquisition of diffusion weighted magnetic resonance images (DW-MRI). While recent advances enable to recover complex fiber geometries using diffusion measurements along various sampling schemes of high order, some older MR systems work with limited gradient tables (ex: maximum of 6 or 12 directions). These systems are designed for Diffusion Tensor Imaging (DTI). Several hospitals and research institutes in the world are limited by these fixed DTI gradient sets. Therefore, groups that want to perform state-of-the-art tractography using high angular resolution diffusion imaging (HARDI) data are penalized and can only perform DTI tractography on their old system. The Gaussian assumption of the tensor model, in DTI, is an over simplification of the diffusion phenomenon of water molecules in the brain and thus cannot resolve crossing fibers. In this work, we show that new diffusion signal modeling and processing techniques enable to capture complex angular structure of the diffusion process even from a reduced gradient direction set arising from an older MR system.

This work has been published in [27].

6.1.1.5. A Robust variational approach for simultaneous smoothing and estimation of DTI

Participants: Rachid Deriche, Meizhu Liu [Department of CISE, University of Florida, Gainesville, USA], Baba C. Vemuri [Department of CISE, University of Florida, Gainesville, USA].

Estimating diffusion tensors is an essential step in many applications — such as diffusion tensor image (DTI) registration, segmentation and fiber tractography. Most of the methods proposed in the literature for this task are not simultaneously statistically robust and feature preserving techniques. In this work, we propose a novel and robust variational framework for simultaneous smoothing and estimation of diffusion tensors from diffusion MRI. Our variational principle makes use of a recently introduced total Kullback–Leibler (tKL) divergence for DTI regularization. tKL is a statistically robust dissimilarity measure for diffusion tensors, and regularization by using tKL ensures the symmetric positive definiteness of tensors automatically. Further, the regularization is weighted by a non-local factor adapted from the conventional non-local means filters. Finally, for the data fidelity, we use the nonlinear least-squares term derived from the Stejskal–Tanner model. We present experimental results depicting the positive performance of our method in comparison to competing methods on synthetic and real data examples.

This work has been accepted for publication in NeuroImage [63].

6.1.2. Modeling in Diffusion MRI

6.1.2.1. Fast and Analytical EAP Approximation from a 4th Order Tensor

Participants: Aurobrata Ghosh, Rachid Deriche.

This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.

Generalized Diffusion Tensor Imaging (GDTI) was developed to model complex Apparent Diffusivity Coefficient (ADC) using Higher Order Tensors (HOT) and to overcome the inherent single-peak shortcoming of DTI. However, the geometry of a complex ADC profile doesn't correspond to the underlying structure of fibers. This tissue geometry can be inferred from the shape of the Ensemble Average Propagator (EAP). Though interesting methods for estimating a positive ADC using 4th order diffusion tensors were developed, GDTI in general was overtaken by other approaches, e.g. the Orientation Distribution Function (ODF), since it is considerably difficult to recuperate the EAP from a HOT model of the ADC in GDTI. In this work, we present a novel closed-form approximation of the EAP using Hermite polynomials from a modified HOT model of the original GDTI-ADC. Since the solution is analytical, it is fast, differentiable, and the approximation converges well to the true EAP. This method also makes the effort of computing a positive ADC worthwhile, since now both the ADC and the EAP can be used and have closed forms. We demonstrate our approach with 4th order tensors on synthetic data and in vivo human data.

This work has been accepted for publication in the International Journal of Biomedical Imaging [54].

6.1.2.2. A Polynomial Approach for Extracting the Extrema of a Spherical Function and its Application in Diffusion MRI

Participants: Aurobrata Ghosh, Elias Tsigaridas [PolSys Project-Team, Inria, Paris Rocquencourt, France], Bernard Mourrain [Galaad Project-Team, Inria, Sophia Antipolis, Méditerranée, France], Rachid Deriche.

This work has been partially supported by the ANR project NucleiPark and the France-Parkinson Association.

Antipodally symmetric spherical functions play a pivotal role in diffusion MRI in representing sub-voxel-resolution microstructural information of the underlying tissue. This information is described by the geometry of the spherical function. In this work, we propose a method to automatically compute all the extrema of a spherical function. We then classify the extrema as maxima, minima and saddle-points to identify the maxima. We take advantage of the fact that a spherical function can be described equivalently in the spherical harmonic (SH) basis, in the symmetric tensor (ST) basis constrained to the sphere, and in the homogeneous polynomial (HP) basis constrained to the sphere. We extract the extrema of the spherical function by computing the stationary points of its constrained HP representation. Instead of using traditional optimization approaches, which are inherently local and require exhaustive search or re-initializations to locate multiple extrema, we use a novel polynomial system solver which analytically brackets all the extrema and refines them numerically, thus missing none and achieving high precision. To illustrate our approach we consider the Orientation Distribution Function (ODF). In diffusion MRI, the ODF is a spherical function which represents a state-of-the-art reconstruction algorithm whose maxima are aligned with the dominant fiber bundles. It is, therefore, vital to correctly compute these maxima to detect the fiber bundle directions. To demonstrate the po- tential of the proposed polynomial approach we compute the extrema of the ODF to extract all its maxima. This polynomial approach is, however, not dependent on the ODF and the framework presented in this work can be applied to any spherical function described in either the SH basis, ST basis or the HP basis.

This work has been submitted to Medical Image Analysis and has been accepted for a publication to appear early 2013 [57].

6.1.2.3. 4th Order Symmetric Tensors and Positive ADC Modelling

Participants: Aurobrata Ghosh, Rachid Deriche.

High Order Cartesian Tensors (HOTs) were introduced in Generalized DTI (GDTI) to overcome the limitations of DTI. HOTs can model the apparent diffusion coefficient (ADC) with greater accuracy than DTI in regions with fiber heterogeneity. Although GDTI HOTs were designed to model positive diffusion, the straightforward least square (LS) estimation of HOTs doesn't guarantee positivity. In this work, we address the problem of estimating 4th order tensors with positive diffusion profiles. Two known methods exist that broach this problem, namely a Riemannian approach based on the algebra of 4th order tensors, and a polynomial approach based on Hilbert's theorem on non-negative ternary quartics. In this work, we review the technicalities of these two approaches, compare them theoretically to show their pros and cons, and compare them against the Euclidean LS estimation on synthetic, phantom and real data to motivate the relevance of the positive diffusion profile constraint.

This work is under submission.

6.1.2.4. Higher-Order Tensors in Diffusion Imaging: A Survey

Participants: Thomas Schultz [MPI for Intelligent Systems, Tubingen, Germany], Andrea Fuster [Eindhoven University of Technology, The Netherlands], Aurobrata Ghosh, Luc Florack [Eindhoven University of Technology, The Netherlands], Rachid Deriche, Lek-Heng Lim [University of Chicago, USA].

Diffusion imaging is a noninvasive tool for probing the microstructure of fibrous nerve and muscle tissue. Higher-order tensors provide a powerful mathematical language to model and analyze the large and complex data that is generated by its modern variants such as High Angular Resolution Diffusion Imaging (HARDI) or Diffusional Kurtosis Imaging. This survey gives a careful introduction to the foundations of higher-order tensor algebra, and explains how some concepts from linear algebra generalize to the higher-order case. From the application side, it reviews a variety of distinct higher-order tensor models that arise in the context of diffusion imaging, such as higher-order diffusion tensors, q-ball or fiber Orientation Distribution Functions (ODFs), and fourth-order covariance and kurtosis tensors. By bridging the gap between mathematical foundations and application, it provides an introduction that is suitable for practitioners and applied mathematicians alike, and propels the field by stimulating further exchange between the two.

This work has been submitted and is under review.

6.1.2.5. Nonnegative Definite EAP and ODF Estimation via a Unified Multi-Shell HARDI Reconstruction

Participants: Rachid Deriche, Jian Cheng [ATHENA and LIAMA, China], Tianzi Jiang [LIAMA, China].

This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.

In High Angular Resolution Diffusion Imaging (HARDI), Orientation Distribution Function (ODF) and Ensemble Average Propagator (EAP) are two important Probability Density Functions (PDFs) which reflect the water diffusion and fiber orientations. Spherical Polar Fourier Imaging (SPFI) is a recent model-free multi-shell HARDI method which estimates both EAP and ODF from the diffusion signals with multiple b values. As physical PDFs, ODFs and EAPs are nonnegative definite respectively in their domains $S2$ and $R3$. However, existing ODF / EAP estimation methods like SPFI seldom consider this natural constraint. Although some works considered the nonnegative constraint on the given discrete samples of ODF / EAP, the estimated ODF/EAP is not guaranteed to be nonnegative definite in the whole continuous domain. The Riemannian framework for ODFs and EAPs has been proposed via the square root parameterization based on pre-estimated ODFs and EAPs by other methods like SPFI. However, there is no work on how to estimate the square root of ODF/EAP called as the wavefuntion directly from diffusion signals. In this work, based on the Riemannian framework for ODFs / EAPs and Spherical Polar Fourier (SPF) basis representation, we propose a unified model-free multi-shell HARDI method, named as Square Root Parameterized Estimation (SRPE), to simultaneously estimate both the wavefunction of EAPs and the nonnegative definite ODFs and EAPs from diffusion signals. The experiments on synthetic data and real data showed SRPE is more robust to noise and has better EAP reconstruction than SPFI, especially for EAP profiles at large radius.

This work has been published in [11] and [18].

6.1.2.6. An Intrinsic Diffeomorphism Invariant Riemannian Framework for Probability Density Function Computing in diffusion MRI

Participants: Rachid Deriche, Jian Cheng [ATHENA and LIAMA, China], Aurobrata Ghosh, Tianzi Jiang [LIAMA, China].

This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.

In High Angular Resolution Imaging (HARDI),Ensemble Average Propagator (EAP) and Orientation Distribution Function (ODF) are two important Probability Density Functions (PDFs), which describe the diffusion probability respectively in 3D space and along directions. Fisher information metric has been successfully applied in Diffusion Tensor Imaging (DTI) on tensor estimation, filtering, registration, statistical analysis, etc. However, to our knowledge, existing works in HARDI mainly focus on ODF/EAP estimation, not on ODF and EAP data processing. In this work, we propose a general state-of-the-art Riemannian framework as a mathematical tool to process such PDF data, by representing the square root of the PDF, called *wavefunction* based on quantum mechanics, as a linear combination of some orthonormal basis functions. The proposed Riemannian framework is showed to be a natural extension of previous Riemannian framework for tensors. We deduced the Riemannian metric for the PDF family via orthonormal basis representation, and proved the statistical manifold to be a convex subset of a high dimensional sphere. In this framework, the exponential map, logarithmic map and geodesic have closed forms, and the weighted Riemannian mean and median uniquely exist. Moreover, we generalized the Log-Euclidean framework and the Geodesic Anisotropy (GA) form tensors to ODFs/EAPs. The theoretical results can be applied to any general PDF data under any orthonormal basis representation. Furthermore we analyzed theoretically the similarities and differences between the Riemannian frameworks for EAPs, ODFs and for tensors, and demonstrated the proposed Riemannian metric is diffeomorphism invariant, which is the natural extension of the previous affine-invariant metric for tensors. Some potential applications were proposed via the Riemannian operations on the ODF/EAP field, such as anisotropy description via GA, nonnegative definite ODF/EAP estimation, interpolation, filtering, Principal Geodesic Analysis (PGA) and atlas estimation. The Riemannian framework and its applications were validated in synthetic, phantom and real data. The experiments demonstrated that the Riemannian framework is very useful for ODF/EAP computing, although the results from Riemannian metric and Euclidean metric are similar for ODFs but much different for EAPs.

This work has been published in [11]. A longer version has been submitted and is under revision for the journal IEEE transaction on Medical Imaging.

6.1.2.7. Ensemble Average Propagator Reconstruction via Compressed Sensing: Discrete or Continuous Bases ?

Participants: Sylvain Merlet, Michael Paquette [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada], Rachid Deriche, Maxime Descoteaux [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada].

This work has been partly supported within the framework of the Brain Connectivities Associate Team.

In this work, we propose to compare the sparsity of two classes of representations for the EAP : The discrete representations, via the Haar, Daubechies-Cohen-Fauveau (DCF) 5-3, DCF 9-7 wavelets bases, and the continuous representations, via Spherical Polar Fourier (SPF) and 3D-SHORE bases.

This work has been published in [28].

6.1.2.8. Parametric dictionary learning for modeling EAP and ODF in diffusion MRI

Participants: Sylvain Merlet, Emmanuel Caruyer, Rachid Deriche.

In this work, we propose an original and efficient approach to exploit the ability of Compressed Sensing (CS) to recover Diffusion MRI (dMRI) signals from a limited number of samples while efficiently recovering important diffusion features such as the Ensemble Average Propagator (EAP) and the Orientation Distribution Function (ODF). Some attempts to sparsely represent the diffusion signal have already been performed. However and contrarly to what has been presented in CS dMRI, in this work we propose and advocate the use of a well adapted learned dictionary and show that it leads to a sparser signal estimation as well as to an efficient reconstruction of very important diffusion features. We first propose to learn and design a sparse and parametric dictionary from a set of training diffusion data. Then, we propose a framework to analytically estimate in closed form two important diffusion features : the EAP and the ODF. Various experiments on synthetic, phantom and human brain data have been carried out and promising results with reduced number of atoms have been obtained on diffusion signal reconstruction, thus illustrating the added value of our method over state-of-the-art SHORE and SPF based approaches.

This work has been published in [25].

6.1.2.9. Constrained Diffusion Kurtosis Imaging Using Ternary Quartics and MLE

Participants: Tristan Milne [Queen's University, Kingston, Ontario, Canada], Aurobrata Ghosh, Rachid Deriche.

This work has been partly supported by the Inria International Internship Program.

We present a ternary quartic based approach with an improved gradient based optimization scheme for diffusion kurtosis imaging to estimate constrained and physically realistic diffusion and kurtosis tensors. We account for the signal noise by considering a maximum likelihood estimation based on the Rician noise model. Diffusion kurtosis imaging (DKI) is a recent important improvement over the diffusion tensor imaging (DTI) model that quantifies the degree of non-Gaussian diffusion in a tissue using a 3D 4th order tensor. However, DKI estimation needs to consider three constraints to be physically relevant. By adopting the implicit ternary quartic parameterization which allows to elegantly impose a positivity constraint on the kurtosis tensor and by employing gradient based optimization schemes, we show dramatically improved performance in terms of estimation time and quality. We derive the mathematical framework and show results on extensive synthetic data experiments.

This work has been published in [30]. A longer version has been submitted and is under revision for the journal Magnetic Resonance in Medicine.

6.1.3. From DW-MRI to Fiber Pathways and Microstructures Recovery

6.1.3.1. From Diffusion MRI to Brain Connectomics

Participants: Aurobrata Ghosh, Rachid Deriche.

Diffusion MRI (dMRI) is a unique modality of MRI which allows one to indirectly examine the microstructure and integrity of the cerebral white matter in vivo and non-invasively. Its success lies in its capacity to reconstruct the axonal connectivity of the neurons, albeit at a coarser resolution, without having to operate on the patient, which can cause radical alterations to the patient's cognition. Thus dMRI is beginning to assume a central role in studying and diagnosing important pathologies of the cerebral white matter, such as Alzheimer's and Parkinson's diseases, as well as in studying its physical structure in vivo. In this work, we present an overview of the mathematical tools that form the framework of dMRI – from modelling the MRI signal and measuring diffusion properties, to reconstructing the axonal connectivity of the cerebral white matter, i.e., from Diffusion Weighted Images (DWIs) to the human connectome.

This work will be published in [55].

6.1.3.2. Biomarkers for HARDI : 2nd & 4th Order Tensor Invariants

Participants: Rachid Deriche, Aurobrata Ghosh, Théodore Papadopoulo.

This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.

In this paper, we explore the theory of tensor invariants as a mathematical framework for computing new biomarkers for HARDI. We present and explain the integrity basis, basic invariants and principal invariants of 2nd & 4th order tensors to expand on a recently proposed paper on 4th order tensor invariants. We present the mathematical results and compute the basic and principal invariants on a controlled synthetic dataset and an in vivo human dataset. We show how the integrity bases of these two sets of invariants can form a promising framework for developing new biomarkers for HARDI.

This work has been published in [22].

6.1.3.3. Generalized Invariants of a 4th order tensor: Building blocks for new biomarkers in dMRI

Participants: Aurobrata Ghosh, Théodore Papadopoulo, Rachid Deriche.

This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.

This paper presents a general and complete (up to degree 4) set of invariants of 3D 4th order tensors with respect to SO3. The invariants to SO3 for the 2nd order diffusion tensor are well known and play a crucial role in deriving important biomarkers for DTI, e.g. MD, FA, RA, etc. But DTI is limited in regions with fiber heterogeneity and DTI biomarkers severely lack specificity. 4th order tensors are both a natural extension to DTI and also form an alternate basis to spherical harmonics for spherical functions. This paper presents a systematic method for computing the SO3 invariants of 3D 4th order tensors, derives relationships between the new (generalized) invariants and existing invariants and shows results on synthetic and real data. It also presents, hitherto unknown, new invariants for 4th order tensors. Analogously to DTI, these new invariants can perhaps form building blocks for new biomarkers.

This work has been published in [23].

6.1.3.4. Tractography via the Ensemble Average Propagator in diffusion MRI

Participants: Sylvain Merlet, Anne-Charlotte Philippe, Rachid Deriche, Maxime Descoteaux [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada].

This work has been partly supported within the framework of the Brain Connectivities Associate Team.

It's well known that in diffusion MRI (dMRI), fibre crossing is an important problem for most existing diffusion tensor imaging (DTI) based tractography algorithms. To overcome these limitations, High Angular Resolution Diffusion Imaging (HARDI) based tractography has been proposed with a particular emphasis on the the Orientation Distribution Function (ODF). In this work, we advocate the use of the Ensemble Average Propagator (EAP) instead of the ODF for tractography in dMRI and propose an original and efficient EAP-based tractography algorithm that outperforms the classical ODF-based tractography, in particular, in the regions that contain complex fibre crossing configurations. Various experimental results including synthetic, phantom and real data illustrate the potential of the approach and clearly show that our method is especially efficient to handle regions where fiber bundles are crossing, and still well handle other fiber bundle configurations such as U-shape and kissing fibers.

This work has been published in [29].

6.1.3.5. Using Radial NMR Profiles to Characterize Pore Size Distributions

Participants: Rachid Deriche, John Treilhard [Queen's University, Ontario, Canada].

This work has been partly supported by the Inria International Internship Program.

Extracting information about axon diameter distributions in the brain is a challenging task which provides useful information for medical purposes; for example, the ability to characterize and monitor axon diameters would be useful in diagnosing and investigating diseases like amyotrophic lateral sclerosis (ALS) or autism. In [78], three families of operators are defined, whose action upon an NMR attenuation signal extracts the moments of the pore size distribution of the ensemble under consideration; also a numerical method is proposed to continuously reconstruct a discretely sampled attenuation profile using the eigenfunctions of the simple harmonic oscillator Hamiltonian – the SHORE basis. The work we have performed here extends this method to other bases that can offer a better description of attenuation signal behaviour – in particular, we proposed the use of the radial Spherical Polar Fourier (SPF) basis. Testing is performed to contrast the efficacy of the radial SPF basis and SHORE basis in practical attenuation signal reconstruction. The robustness of the method to additive noise is tested and analyzed. We demonstrated that a low-order attenuation signal reconstruction outperforms a higher-order reconstruction in subsequent moment estimation under noisy conditions. We proposed the simulated annealing algorithm for basis function scale parameter estimation. Finally, analytic expressions are derived and presented for the action of the operators on the radial SPF basis (obviating the need for numerical integration, thus avoiding a spectrum of possible sources of error).

This work has been published [20].

6.1.3.6. Elliptic Fourier Features of Brain White Matter Pathways

Participants: Rachid Deriche, Ali Demir [Sabancy University, TU], Gozde Unal [Sabancy University, TU].

Magnetic resonance imaging provides diffusion weighted images (DMRI), which non-invasively reconstruct the brain white matter pathways. DMRI is used to study brain white matter diseases as well as aid surgical planning. As localization of different white matter pathways surrounding a pathology is crucial for surgical planning, automatic extraction and classification of different anatomical white matter pathways pre-operatively becomes an important computational tool. In this work, we propose a method for classification of brain white matter pathways based on 3D elliptic Fourier descriptors, which are extended from the 2D elliptic Fourier descriptors. We performed experiments and validation of the proposed method on a white matter atlas space and on real pathological cases.

This work has been published [41].

6.2. Multi-Imaging Modalities

6.2.1. Coupling functional and structural models

6.2.1.1. A nested cortex parcellation combining analysis of MEG forward problem and diffusion MRI tractography

Participants: Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

Understanding the relationship between structure and function is a major challenge in neuroscience. Diffusion MRI (dMRI) in the only non-invasive modality allowing to have access to the neural structure. Magnetoencephalography (MEG) is another non-invasive modality that allows a direct access to the temporal succession of cognitive processes. Functional cortex parcellation being one of the most important ways to understanding structure-function relationship, we propose an innovative method merging MEG and dMRI to parcellate the cortex. The combination of MEG forward problem and connectivity information reveals cortical areas generating a similar magnetic field at sensors while having a similar connectivity. Results show suitable clusters that forecast interesting studies for inter- and intra- subjects comparisons of the cortex parcellations. The automatic nested cortex parcellation we propose could be a first step to analyse sources that are seeds of long or short range connectivity and to differentiate these connectivities in the white matter

This work has been published in [31].

6.2.1.2. dMRI tractography of WM fibers to recover the anatomical connectivity supporting a MEG epileptic network

Participants: Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

Cerebral organization is determined by segregated and integrated regions both functionally and anatomically. These cerebral networks are the foundations of the execution of major part of cognitive processes. Information about the structure of the white matter (WM) and the functionality of networks are both needed to understand these cerebral networks.

This work proposes an efficient method to inform a given functional network on its anatomical support: how many anatomical connections exist between functionally connected regions and what are their geometries . Diffusion MRI being the only non invasive method allowing to have access to the micro-structure of the WM, we used diffusion information to underline the degree of connectivity between functionally connected regions while taking advantage of WM fibers reconstruction to determine the way taken by the anatomical network supporting the functional network.

Due to the complex dynamical alteration of epilepsy, the study of large-scale functional connectivity is difficult. But diffusion imaging studies have shown alterations of the WM between epileptic zones and connected areas. This methodology allows to add qualitative (degree of connectivity) and geometrical (WM fibers reconstruction) information on the anatomical network supporting an epileptic network mostly determined by magneto-encephalography (MEG).

This work has been published in [35].

6.2.1.3. Whole cortex parcellation combining analysis of MEG forward problem, structural connectivity and Brodmann's atlas

Participants: Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

Functional cortex parcellation is one of the most important ways to understand the link between structure and function in the brain. Brodmann's atlas remains a fundamental pillar to understand this relationship because its areas are defined by similar cytoarchitecture and functional imaging notably had revealed that they correspond, entirely or in part, to functional areas. So, its integration to diffusion MRI (dMRI) data is pertinent, dMRI being the only non invasive and in-vivo imaging modality able to have access to a detailed geometric description of the anatomical connectivity between brain areas. In this work, we propose to define a new connectivity profile of cortical sources based on the Brodmann's atlas. After its registration to T1 and diffusion weighted images of the same subject, we reconstructed the brain surfaces and considered the cortical sources to be the vertices of the white matter/ grey matter boundary mesh. We performed a probabilistic tractography taking each cortical sources as seeds and the L Brodmann's areas as L targets. Thus, we obtained the connectivity profile of a cortical source: a vector v of size L where v (l) is the degree of connectivity of the source to the l th Brodmann's area. Then, we developped a cortical parcellation method jointly analyzing the MEG forward problem and the connectivity profiles based on Brodmann's atlas of cortical sources. We computed the leadfield matrix that relates the sources to the MEG sensors. We applied a k-means algorithm to the leadfield matrix to cluster sources having a close magnetic field to the MEG sensors. Then, in each leadfield-based cluster, we clustered sources via their connectivity profile based on Brodmann's atlas. This automatic parcellation is an efficient preprocessing to compute a MEG inverse problem on functional data informed by its structural connectivity.

This work has been published in [32].

6.2.1.4. Study of the brain connectivity in an Immersive Space

Participants: Anne-Charlotte Philippe, Jean-Christophe Lombardo [Dream Project-Team, Inria, Sophia Antipolis, Méditerranée, France].

Virtual reality is a powerful tool for scientific visualization. When the amount and complexity of the visualized data grows, standard visualization applications on desktop computers become inefficient. In this work, we present the use of a CAVE like VR facility in a neuroscientific context. The aim is to have a better understanding of the brain connectivity. Both anatomical and functional data are attached to a mesh representing the brain surface.

Specific tools developed for this study and the way we used them are presented in [36] emphasizing drawbacks and advantages of virtual reality in a scientific visualization context.

This work has been published in [36].

6.2.1.5. Cortex parcellation via diffusion data as prior knowledge for the MEG inverse problem

Participants: Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

In this work, we present a new approach to the recovery of dipole magnitudes in a distributed source mo-del for magnetoencephalographic (MEG) imaging. This method consists in introducing prior knowledge regarding the anatomical connectivity in the brain to this ill-posed inverse problem. Thus, we perform cortex parcellation via structural information coming from diffusion MRI (dMRI), the only non-invasive modality allowing to have access to the structure of the WM tissues. Then, we constrain, in the MEG inverse problem, sources in the same diffusion parcel to have close magnitude values. Results of our method on MEG simulations are presented and favorably compared with classical source reconstruction methods.

This work is currently under submission.

6.2.1.6. Fractality in the neuron axonal topography of the human brain based on 3-D diffusion MRI

Participants: Panayotis Katsaloulis [Institute of Physical Chemistry "Demokritos" (IPC),National Center for Scientific Research "Demokritos", Greece], Aurobrata Ghosh, Anne-Charlotte Philippe, Astero Provata [Institute of Physical Chemistry "Demokritos" (IPC),National Center for Scientific Research "Demokritos", Greece], Rachid Deriche.

In this work, we conduct a group study, with 18 subjects, to validate the computational robustness of the fractal dimension of the neuron axonal topography in the human brain that is derived from diffusion MRI (dMRI) acquisitions. We extend the work done in a previous paper by some of the current authors where the fractal dimension of the neuron axonal topography from dMRI data was computed from 2-D regions of interest. The fractal dimensions Df of the entire 3-D volume of the brain is here estimated via the Box Counting, the Correlation DImension and the Fractal Mass Dimension methods. 3-D neuron axon data are obtained using tractography algorithms on Diffusion Tensor Imaging of the brain. We find that all three calculations of Df give consistent results across subjects, namely, they demonstrate fractal characteristics in the short and medium length scales: different fractal exponents prevail at different length scales, an indication of multifractality. We surmise that this complexity stems as a collective property emerging when many local brain units performing different functional tasks and having different local topologies are recorded together.

This work has been published in [15].

6.3. Forward and Inverse Problems

6.3.1. Source localization using rational approximation on plane sections

Participants: Maureen Clerc, Théodore Papadopoulo, Juliette Leblond [Apics Project-Team, Inria, Sophia Antipolis, Méditerranée, France], Jean-Paul Marmorat [CMA, Ecole des Mines Paristech, Sophia Antipolis, France].

In functional neuroimaging, a crucial problem is to localize active sources within the brain non-invasively, from knowledge of electromagnetic measurements outside the head. Identification of point sources from boundary measurements is an ill-posed inverse problem. In the case of electroencephalography (EEG), measurements are only available at electrode positions, the number of sources is not known in advance and the medium within the head is inhomogeneous. This work presents a new method for EEG source localization, based on rational approximation techniques in the complex plane. The method is used in the context of a nested sphere head model, in combination with a cortical mapping procedure. Results on simulated data prove the applicability of the method in the context of realistic measurement configurations.

This work has been published in the journal Inverse Problems [14].

6.3.2. The adjoint method of OpenMEEG for EEG and MEG with large source space

Participants: Maureen Clerc, Théodore Papadopoulo, Alexandre Gramfort [Telecom Paristech], Emmanuel Olivi [Forner member of the Athena Project-Team].

In EEG or MEG, a lead field is the linear operator which associates unitary dipolar sources to the resulting set of sensor measurements. In practise, the source space often includes over 10 000 dipoles, which sometimes causes memory problems. The adjoint approach considers the forward problem from the viewpoint of sensors instead of sources: this drops down the number of linear systems to solve by two orders of magnitude. The adjoint approach is here proposed in the context of the Boundary Element Method, and its implementation is provided by the OpenMEEG library.

This work was presented at the BIOMAG conference [38].

6.3.3. Comparison of Boundary Element and Finite Element Approaches to the EEG Forward Problem

Participants: Maureen Clerc, Carsten Wolters [Institute for Biomagnetism and Biosignal Analysis, University of Münster], Johannes Vorwerk [Institute for Biomagnetism and Biosignal Analysis, University of Münster], Martin Burger [Institut für Numerische und Angewandte Mathematik, Fachbereich Mathematik und Informatik, Westfälische Wilhelms Universität (WWU) Münster], Jan de Munck [Vrije Universiteit Medical Centre (VUMC), The Netherlands].

The accurate simulation of the electric fields evoked by neural activity is crucial for solving the inverse problem of EEG. Nowadays, boundary element methods (BEM) are frequently applied to achieve this goal, usually relying on the simpli-fication of approximating the human head by three nested compartments with isotropic conductivities (skin, skull, brain). Here, including the highly-conducting cerebrospinal fluid (CSF) is a difficult task due to the complex geometrical structure of the CSF, demanding a high number of additional nodes for an accurate modeling and thus a strongly increased computational effort. Though, CSF conductivity is well-known and nearly not varying inter-individually and its significant influence on EEG forward simulation has been shown. The CSF can be included at negligible computational costs when applying finite element (FE) forward approaches. In this study we compare the accuracy and performance of state-of-the-art BE and FE approaches in both artificial and realistic three layer head models, showing that all approaches lead to high numerical accuracies. Furthermore, we demonstrate the significant influence of modeling the CSF compartment as disregarding this compartment leads to model errors that lie clearly above the observed numerical errors.

A book chapter on BEM and FEM models has been published in the Handbook for Neural Activity Measurement [40]. The comparison was presented at the BIOMAG conference [19].

6.3.4. Domain Decomposition to handle versatile conductivity models

Participants: Maureen Clerc, Théodore Papadopoulo, Emmanuel Olivi [Forner member of the Athena Project-Team].

Source localization from external data such EEG or MEG, requires a good understanding of the electromagnetic behavior of the patient head. Several models can been used, representing more or less complex geometrical shapes, and conductivity profiles. Different numerical methods allow to cope with different types of models: the Finite Element Method (FEM) can handle very general conductivity models, whereas the Boundary Element Method (BEM) is limited to piecewise constant conductivity. On the other hand, it is easier with BEM than with FEM to accurately represent sources in isotropic media. Thanks to domain decomposition, we propose to solve a EEG forward problem using BEM where the sources are (the brain) and FEM for other tissues (with notably inhomogeneities in the skull).

This work was presented at the BIOMAG conference [37].

6.4. Brain Computer Interfaces

6.4.1. Combining ERD and ERS features to create a system-paced BCI

Participants: Maureen Clerc, Théodore Papadopoulo, Joan Fruitet, Eoin Thomas.

An important factor in the usability of a brain computer interface (BCI) is the setup and calibration time required for the interface to function accurately. Recently, brain-switches based on the rebound following motor imagery of a single limb effector have been investigated as basic BCIs due to their good performance with limited electrodes, and brief training session requirements. Here, a BCI is proposed which expands the methodology of brain-switches to design an interface composed of multiple brain-buttons. The algorithm is designed as a system paced interface which can recognise 2 intentional-control tasks and a no-control state based on the activity during and following motor imagery in only 3 electroencephalogram channels. An online experiment was performed over 6 subjects to validate the algorithm, and the results show that a working BCI can be trained from a single calibration session and that the post motor imagery features are both informative and robust over multiple sessions.

This work, which was partially presented at the EMBS conference [33], is currently under revision for the Journal of Neuroscience Methods.

6.4.2. Bandit algorithms for faster task selection in BCI

Participants: Maureen Clerc, Joan Fruitet, Alexandra Carpentier [Sequel Project-Team, Inria Lille, France], Rémi Munos [Sequel Project-Team, Inria Lille, France].

BCIs based on sensorimotor rhythms use a variety of motor tasks, such as imagining moving the right or left hand, the feet or the tongue. Finding the tasks that yield best performance, specifically to each user, is a time consuming preliminary phase to a BCI experiment. This study presents a new adaptive procedure to automatically select, online, the most promising motor task for an asynchronous brain-controlled button.
We develop for this purpose an adaptive *Upper Confidence Bound* algorithm based on the stochastic bandit theory, and design an EEG experiment to test our method. We compare (offline) the adaptive algorithm to a naive selection strategy which uses uniformly distributed samples from each task. We also run the adaptive algorithm online to fully validate the approach.
By not wasting time on inefficient tasks, and focusing on the most promising ones, this algorithm results in a faster task selection and a more efficient use of the BCI training session. More precisely, the offline analysis reveals that the use of this algorithm can reduce the time needed to select the most appropriate task by almost half without loss in precision, or alternatively, allow to investigate twice the number of tasks within a similar time span. Online tests confirm that the method leads to an optimal task selection.
This study is the first one to optimize the task selection phase by an adaptive procedure. By increasing the number of tasks that can be tested in a given time span, the proposed method could contribute to reducing “BCI illiteracy”.

This work is the result of the collaboration between Sequel and Athena within the ANR CoAdapt. It has been published in NIPS [21] and is accepted in the Journal of Neural Engineering [52].

6.4.3. An analysis of performance evaluation for motor-imagery based BCI

Participants: Maureen Clerc, Matthew Dyson [Laboratoire de Neurosciences Cognitives, Aix-Marseille Université, France], Eoin Thomas.

In recent years, numerous brain computer interfaces (BCIs) have been proposed which incorporate features such as adaptive classification, error detection and correction, fusion with auxiliary signals and shared control capabilities. Due to the added complexity of such algorithms, the evaluation strategy and metrics used for analysis must be carefully chosen to accurately represent the performance of the BCI. In this article, metrics are reviewed and contrasted using both simulated examples and experimental data. Furthermore, a review of the recent literature is presented to determine how BCIs are evaluated, in particular focusing on the correlation between how the data are used relative to the BCI subcomponent under investigation. From the analysis performed in this study, valuable guidelines are presented regarding the choice of metrics and evaluation strategy dependent upon any chosen BCI paradigm.

This work was supported by the ANR Co-Adapt and is currently under revision for the Journal of Neural Engineering.

BAMBOO Project-Team

6. New Results

6.1. Partial Enumeration of Traces

Traditional algorithms to solve the problem of sorting by signed reversals output just one optimal solution while the space of all optimal solutions can be huge. A so-called trace represents a group of solutions which share the same set of reversals that must be applied to sort the original permutation following a partial ordering. By using traces, we therefore can represent the set of optimal solutions in a more compact way. Algorithms for enumerating the complete set of traces of solutions were developed. However, due to their exponential complexity, their practical use is limited to small permutations. A partial enumeration of traces is a sampling of the complete set of traces and can be an alternative for the study of distinct evolutionary scenarios of big permutations. Ideally, the sampling should be done uniformly from the space of all optimal solutions. This is however conjectured to be #P-complete.

We proposed and evaluated three algorithms for producing a sampling of the complete set of traces that instead can be shown in practice to preserve some of the characteristics of the space of all solutions [7]. We analysed the distribution of the enumerated traces with respect to their height and average reversal length.

6.2. De-novo calling alternative splicing events from RNA-seq data

We addressed the problem of identifying and quantifying polymorphisms in RNA-seq data when no reference genome is available, without assembling the full transcripts. Based on the fundamental idea that each polymorphism corresponds to a recognisable pattern in a De Bruijn graph constructed from the RNA-seq reads, we proposed a general model for all polymorphisms in such graphs. We then introduced an exact algorithm, called KISSPLICE, to extract alternative splicing events. The first version of KISSPLICE appeared in 2011, but several important improvements were implemented in 2012 [24]. The first improvement was the memory consumption, the new version is much more memory efficient and can handle datasets of approximately 10^8 reads. The second was in the running time, the enumeration step can now be done in parallel, which results in a significant speedup in the overall running time. Finally, an improved event quantification step was added to the method.

Application-wise, we showed that KISSPLICE enables to identify more correct events than general purpose transcriptome assemblers. Additionally, on a 71 M reads dataset from human brain and liver tissues, KISSPLICE identified 3497 alternative splicing events, out of which 56% are not present in the annotations, which confirms recent estimates showing that the complexity of alternative splicing has been largely underestimated so far.

6.3. Efficient bubble and/or cycle enumeration in directed/undirected graphs

Polymorphisms in DNA- or RNA-seq data lead to recognisable patterns in a de Bruijn graph representation of the reads obtained by sequencing. Such patterns have been called mouths, or bubbles in the literature. They correspond to two vertex-disjoint directed paths between a source s and a target t. Due to the high number of such bubbles that may be present in real data, their enumeration is a major issue concerning the efficiency of dedicated algorithms. We developed the first linear delay algorithm to enumerate all bubbles with a given source [31].

By combining the insights from the most efficient but not optimal solution presented by Johnson [SIAM J. Computing, 1975] for simple cycle enumeration in undirected graphs and an amortisation technique previously established by our collaborators Roberto Grossi and Rui Ferreira [ESA, 2011] from the University of Pisa, Italy, we obtained the first optimal solution to list all the simple cycles in an undirected graph G (paper accepted at SODA 2013, to appear). Moreover, we also obtained the first optimal solution to list all the simple paths from s to t in an undirected graph G. This work benefited also from discussions and work from Pierluigi Crescenzi and Marie-France Sagot. The method is not naturally extendable to directed graphs, and the challenge is now to obtain optimal solutions in this case also.

6.4. Simulating RNA-seq experiments

RNAseq experiments now enable to characterise the RNA complement of a cell. However, the series of steps (fragmentation, reverse transcription, sequencing) that separate the initial RNA molecules from the short DNA reads obtained in fine are not well understood although it is widely accepted that they contribute to generating noise in the signal. We introduced the FLUXSIMULATOR [14], a computer program able to reproduce the biases seen in RNAseq data. This pipeline should prove useful both to produce realistic data on which to test programs which aim at reconstructing RNA from short reads, and suggest ways of improving the experimental steps so that they produce less noise.

6.5. Chimeric Transcripts may be Translated

There is now increasing evidence for the existence of so-called Chimeric Transcripts. In contrast to regular transcripts, which are composed of exons located close to each other on the genome, these chimeric transcripts can be composed of exons which are located megabases away, or even on different chromosomes. We showed that these chimeras are lowly expressed, are tissue specific, and that some of them may be translated, yielding proteins with altered function or localisation [13].

6.6. Transcriptomics of symbiosis in the Asobara tabida-Wolbachia association

Wolbachia has evolved a very peculiar phenotype in the host *Asobara tabida* where it is obligatory for oogenesis. Transcriptomics approaches were developed first using Sanger sequencing of mRNA [19]. It has now been complemented by RNAseq analyses on two lines, which exhibit different ovarian phenotypes in absence of *Wolbachia*. We have currently analysed these data both to isolate genes that are differentially expressed, but also that exhibit polymorphism between the lines. Interesting candidates were detected that are under further investigation and that are involved in the regulation of early oogenesis, apoptosis, autophagy and oxidative stress. This part is in direct connection with the algorithms developed by BAMBOO for the analysis of NGS data without a reference genome (KISSPLICE).

6.7. MicroRNA predictor

We developed a microRNA predictor using structural and target information. The method shows 97% sensitivity and 90% specificity for the *Acyrthosiphon pisum* genome. Comparing to the results of the previous method we developed in 2010 (available in the software CRAVELA) we obtained a better performance (sensitivity 90% and specificity 88%). However, as we are working on a genome wide scale, it is important to obtain even better specificity (obviously, maintaining a reasonable sensitivity). This work is currently in development.

On the other hand, the computational search for novel miRNA precursors often involves also some sort of structural analysis with the aim of identifying which type of structures are recognised and processed by the cellular miRNA-maturation machinery. A natural way to tackle this problem is to perform clustering over the candidate structures along with known miRNA precursor structures. Mixed clusters allows then the identification of candidates that are similar to known precursors. Given the large number of pre-miRNA candidates that can be identified in single-genome approaches, even after applying several filters for precursor robustness and stability, a conventional structural clustering approach is unfeasible. We proposed a method to represent candidate structures in a feature space which summarises key sequence/structure characteristics of each candidate [21]. We showed that proximity in this feature space is related to sequence/structure similarity, and we selected candidates which have a high similarity to known precursors. Additional filtering steps were then applied to further reduce the number of candidates to those with greater transcriptional potential.

6.8. Genomics of symbiosis

Insect symbioses are model systems for studying the effect of symbionts and the evolution of bacterial genomes. Members from the LBBE described the symbiotic complement of different biotypes of the insect *Bemisia tabaci* in Western Africa. We further obtained the complete genome of different symbionts that co-exist in *Bemisia tabaci*, among which the the primary symbiont *Portiera* [25], *Hamiltonella*, *Rickettsia* and *Wolbachia*. Analyses are underway, that concern the possible complementation between *Hamiltonella* and *Portiera* and the comparative analyses of different *Hamiltonella* genomes.

6.9. Representation and curation of metabolic data: UniPathway, Rhea and MNX

These activities are carried out in collaboration with the SwissProt group at the Swiss Institute for BioInformatics (SIB). UNIPATHWAY (http://www.unipathway.org) is a manually curated database of metabolic pathways. It provides the official controlled vocabulary for pathway annotation within UNIPROTKB records since 2009. A complete description of the UNIPATHWAY database and of its relationship with UNIPROTKB has been published in *Nucleic Acids Research* (Jan. 2012 Database Issue) [22]. RHEA (http://www.ebi.ac.uk/rhea) is developed jointly with the European Institute for Bioinformatics (EBI) and the SIB. It provides a comprehensive resource of expert-curated biochemical reactions, for use in a large spectrum of applications, including metabolic network reconstruction and pathway inference. The complete description of the RHEA database appeared in the Jan. 2012 *NAR* Database issue [5]. The MNX project is developed in the context of the METANETX project (http://www.metanetx.org). It attempts to automate the reconciliation of discrepancies between metabolite or reaction information from distinct resources (BIGG, BRENDA, CHEBI/RHEA, KEGG, METACYC, UNIPATHWAY, THE SEED, REACTOME), thereby alleviating a major bottleneck in the construction of genome-scale metabolic network models. The MNXREF namespace is available at http://www.metanetx.org/mnxdoc/mnxref.html and the method to compute the MNXREF namespace is described in [8].

6.10. Annotation of the proteins of Angomonas deanei and Strigomonas culicis

Angomonas deanei and *Strigomonas culicis* are trypanosomatids that harbour only one beta-proteobacterial endosymbiont and this mutualistic association is an interesting model to study eukaryotic cell evolution. The genomes of these organisms were sequenced by our collaborators at LNCC / MCT (Brazil) and we participated in the functional annotation of these genomes as concerns their metabolism which enabled to reveal new aspects of the *Trypanosomatidae* family. This work has been submitted for publication. It was done with Ana Tereza Vasconcelos in a collaboration with Maria Cristina Machado Motta (UFRJ - Brazil).

6.11. Finding candidate genes for orphan enzymes

Of all biochemically characterized metabolic reactions formalized by the IUBMB, over one out of four have yet to be associated with a nucleic or protein sequence, *i.e.* are sequence-orphan enzymatic activities. Few bioinformatics annotation tools are able to propose candidate genes for such activities by exploiting context-dependent rather than sequence-dependent data, and none are readily accessible and propose result integration across multiple genomes. We introduced CANOE (Candidate genes for Orphan Enzymes), a four-step bioinformatics strategy that proposes ranked candidate genes for sequence-orphan enzymatic activities (or orphan enzymes for short) [26]. Our strategy found over 60,000 genomic metabolons in more than 1,000 prokaryote organisms from the MICROSCOPE platform developed by the group of Claudine Médigue from the Génoscope with whom this work was done, generating candidate genes for many metabolic reactions, of which more than 70 distinct orphan reactions. A computational validation of the approach was discussed and we presented a case study on the anaerobic allantoin degradation pathway in *Escherichia coli* K-12.

6.12. Metabolic cooperation of symbionts and their host trypanosomatids

Trypanosomatids that harbour a symbiotic bacterium (SHTs) are known to have less nutritional requirements when compared to their counterparts without symbionts (RTs). Nutritional and biochemical data indicated that the symbionts largely contributed to the routes for amino acid and vitamin biosynthesis. We analysed the genomic data of 5 SHTs and their respective symbionts and 2 RTs and we found most of the genes related to those pathways in the symbionts. This work will soon be submitted for publication. It is being done with Ana Tereza Vasconcelos in a collaboration with Erney P. Camargo, Marta M.G. Teixeira (USP - Brazil), João M.P. Alves, Gregory A. Buck (VCU - USA), and Maria Cristina Machado Motta (UFRJ - Brazil).

6.13. Structural and dynamical analysis of biological networks

We published a review on the structural and dynamical analysis of biological networks with as main focus explaining the cares that should be taken when this kind of analysis is performed [18]. Correctly distinguishing between potential metabolic networks and their realisations is necessary in choosing the right methods to be used and in the interpretation of their outcomes. In our review, we covered several different techniques, both static and dynamic, for the analysis of metabolic networks such as centrality techniques, flux-balance analysis and kinetic modelling of full-scale networks.

6.14. Network distance analysis

We addressed the diameter computation problem in the case of undirected unweighted graphs, where the diameter D is defined as the maximum distance among all the pairs of nodes and the distance $d(u, v)$ between two nodes u and v is defined as the number of edges contained in the shortest path from u to v. In the context of real-world networks, the textbook method based on performing a breadth-first search (in short, BFS) from *every* node of the graph, requires a prohibitive cost of $O(nm)$ time, where n is the number of nodes and m is the number of edges of the graph. Our main contribution consists of showing that BFS can indeed be an extremely *powerful* tool to compute the *exact* value of the diameter, whenever it is used in a more clever way. In particular, we have developed the *iterative* Fringe Upper Bound (in short, iFUB) algorithm to calculate the exact value of the diameter. This work has been accepted for publication in *Theoretical Computer Science* (to appear).

We then successively generalised the idea of the iFUB algorithm, by presenting the directed iFUB (in short, DiFUB) algorithm, in order to calculate the diameter of the strongly connected components of directed graphs [33]. As far as we know, DiFUB is the first algorithm which is able to compute exactly the diameter of the strongly connected components of huge real-world directed graphs. The DiFUB algorithm can also return a pair of nodes whose distance is exactly equal to the diameter, and a natural adaptation of it works also for weighted graphs.

6.15. Information spreading in dynamic graphs

We showed how a technique used to analyse the flooding completion time in the case of a special class of random evolving graph model, that is, the *edge-Markovian model*, can be used in order to prove that the flooding completion time of a random evolving graph $(G_t)_{t\geq 0}$ is bounded by $kD + 2C$, where intuitively (1) k is the smallest time necessary for the rising of a giant component, (2) D is the diameter of the giant component, and (3) C is the time required for the nodes outside the giant component to eventually get an edge connecting them to the giant component [30]. Then, based on this result, we developed a general methodology for analysing flooding in sequences of random graphs and we applied this general methodology to the case of power-law evolving graphs (that is, sequences of mutually independent random graphs such that the number y of nodes of degree x distributes like $1/x^\beta$ for some $\beta > 0$), and to the case of an arbitrary given degree distribution.

6.16. Metabolic network comparison

Previous works on minimal gene sets, when analysing host-dependent bacteria, found small common sets of metabolic genes. When such analyses are restricted to bacteria with similar lifestyles, larger portions of metabolism are expected to be shared and their composition is worth investigating. Comparing the small molecule metabolism of 58 bacteria carefully selected and representing a range of lifestyles, we found not a single enzymatic reaction common to all of them. While obligate intracellular symbionts have no core of reactions within their group, extracellular and cell-associated symbionts do have a small core enriched in biosynthetic processes composed of disconnected fragments. As more genomes are added, we expect, based on our simulations, that the core of cell-associated and extracellular bacteria continues to diminish, converging to approximately 60 reactions. These results were in preparation in 2011 and are now published [17]. The work was done with Ana Tereza Vasconcelos and in a collaboration with Ludovic Cottret (INSA Toulouse).

6.17. Core and periphery of metabolic networks

The core metabolism can be defined as the reactions present in every organism, however it is not robust considering that adding or removing one organism in the study will modify the resulting set. An alternative way is to include in the core the reaction that is present in a large enough proportion of species. For that, we proposed a method where the threshold to decide what is large enough is not set by the user (thus relying on a subjective choice), but rather automatically selected by the method, relying on the information contained in the data. Two approaches are being proposed, one is EM (Expectation Maximization) which relies only on the information of presence / absence of a reaction in a species while the second (NEM - Neighboring Expectation Maximization) relies on a neighbouring relation between reactions. The latter tends to classify in a same group (core or periphery) a reaction for which a majority of neighbours belong to a same group. The work is being done with Ana Tereza Vasconcelos in a collaboration with Catherine Matias, Christophe Ambroise, Yolande Diaz (Genopole, CNRS).

6.18. Metabolic stories

Enumerating stories, *i.e.*, enumerating maximal directed acyclic graphs with sets of sources and targets contained in a given subset of the nodes, is an algorithmic approach we proposed for interpreting metabolomics experiments. The modelling, algorithms and complexity results were recently accepted for publication [2]. The complexity of the enumeration problem remains unknown. There are also further modelling issues that could be dealt with in a near future. Both considerations were also detailed in a talk given in August at St. Petersburg, in the First RECOMB Satellite Conference on Open Problems in Algorithmic Biology.

We then applied our enumerating method on real data. We analysed data on the detoxification process of yeast cells exposed to cadmium. Our method allowed to recover known pathways involved in the process but also to propose alternative scenarios. The method was also investigated in order to automatically propose metabolic pathways through an experiment in which we try to recover known metabolic pathways using only minimal information (*e.g.*, their entries and endpoints). A paper is in prepartion and should soon be submitted for publication. This work is being done in collaboration with Fabien Jourdan and Ludovic Cottret from the INRA at Toulouse, and with Christophe Junot from the CEA in Paris.

6.19. Minimal precursor sets

We proposed two new, more efficient algorithms for the enumeration of minimal precursor sets: PITUFINA and PAPA PITUFO [3]. The model of minimal precursor sets we had previously published was the first to formally take into account cycles, which are a common event in metabolic networks. The new methods avoid the memory issues of our previous approach by traversing directly the metabolic network structure instead of building a secondary tree representation. PAPA PITUFO additionally saves pre-computed solutions by a local modification of the network.

6.20. Minimum ratio cover of matrix columns by extreme rays of its induced cone

Given a matrix S and a subset of columns R, we studied the problem of finding a cover of R with extreme rays of the cone $\mathcal{F} = \{v \in \mathbb{R}^n \mid Sv = \mathbf{0}, v \geq \mathbf{0}\}$, where an extreme ray v covers a column k if $v_k > 0$ [34]. In order to measure how proportional a cover is, we introduced two different minimisation problems, namely the MINIMUM GLOBAL RATIO COVER (MGRC) and the MINIMUM LOCAL RATIO COVER (MLRC) problems. In both cases, we applied the notion of the *ratio* of a vector v, which is given by $\frac{\max_i v_i}{\min_{j|v_j>0} v_j}$. These problems were originally motivated by a biological question on metabolic networks. This notion of ratio is also of interest in the field of *exact linear programming*, where current algorithms for scaling a matrix have a complexity that depends on the ratio of its elements. We showed that these two problems are NP-hard, even in the case in which $|R| = 1$. We introduced a mixed integer programming formulation for the MGRC problem, which is solvable in polynomial time if all columns should be covered, and introduce a branch-and-cut algorithm for the MLRC problem. Finally, we presented computational experiments on data obtained from real metabolic networks.

6.21. Optimal flux spaces of genome-scale stoichiometric models

The metabolism of organisms can be studied with comprehensive stoichiometric models of their metabolic networks. Flux balance analysis (FBA) calculates optimal metabolic performance of stoichiometric models. However, detailed biological interpretation of FBA is limited because, in general, a huge number of flux patterns give rise to the same optimal performance. The complete description of the resulting optimal solution spaces was thus far a computationally intractable problem. We introduced COPE-FBA: Comprehensive Polyhedra Enumeration Flux Balance Analysis, a computational method that solves this problem [15]. COPE-FBA indicates that the thousands to millions of optimal flux patterns result from a combinatorial explosion of flux patterns in just a few metabolic sub-networks. The entire optimal solution space can now be compactly described in terms of the topology of these sub-networks. COPE-FBA simplifies the biological interpretation of stoichiometric models of metabolism, and provides a profound understanding of metabolic flexibility in optimal states.

6.22. Lateral gene transfer as a support for the tree of life

We published with Sophie Abby the last results of her PhD work that apply an explicit phylogenetic model of horizontal gene transfer to bacterial and archaeal phyla [1]. We showed that lateral gene transfer allows to discriminate between phylogenetic hypotheses, and that in a typical bacterial gene family, 96-98% of tree branches result from vertical descent and 2-4% from lateral gene transfer.

6.23. Comparative approximability of hybridization number and directed feedback vertex set

We showed that the problem of computing the hybridization number of two rooted binary phylogenetic trees on the same set of taxa X has a constant factor polynomial-time approximation if and only if the problem of computing a minimum-size feedback vertex set in a directed graph (DFVS) has a constant factor polynomial-time approximation. The latter problem, which asks for a minimum number of vertices to be removed from a directed graph to transform it into a directed acyclic graph, is one of the problems in Karp's seminal 1972 list of 21 NP-complete problems. However, despite considerable attention from the combinatorial optimisation community, it remains to this day unknown whether a constant factor polynomial-time approximation exists for DFVS. Our result thus placed the (in)approximability of hybridization number in a much broader complexity context, and as a consequence we obtained that hybridization number inherits inapproximability results from the problem Vertex Cover [16]. On the positive side, we used results from the DFVS literature to give an $O(logrloglogr)$ approximation for the hybridization number, where r is the value of an optimal solution to the hybridization number problem. This work is submitted for publication.

6.24. Influence of symbionts on antagonistic interactions

Symbionts are often key players in antagonistic interactions between their hosts and other organisms. In host-parasitoid interactions, both players can be infected by different symbionts. We investigated how a virus and *Wolbachia*, respectively infecting a parasitoid and a drosophila, can shape the host-parasitoid interaction. While only a limited effect *Wolbachia* has been detected, the virus protects the parasitoid from the immune response of *Drosophila* [20]. Protection conferred by symbionts to their insect hosts is a promising avenue for antivectorial programs, but requires a thorough analysis of the evolutionary consequences of protection. We reviewed the literature on this topic [28].

6.25. Mod/Resc Parsimony Inference

We addressed a computational biology problem that aims at understanding a mechanism that could potentially be used to genetically manipulate natural insect populations infected by inherited, intra-cellular parasitic bacteria. In this problem, that we denoted by Mod/Resc Parsimony Inference, we are given a boolean matrix and the goal is to find two other boolean matrices with a minimum number of columns such that an appropriately defined operation on these matrices gives back the input. We showed that this is formally equivalent to the Biclique Edge Cover for Bipartite Graphs problem and derive some complexity results for our problem using this equivalence. We provided a new, fixed parameter tractability approach for solving both problems that slightly improves upon a previously published algorithm for the Biclique Edge Cover for Bipartite Graphs. Finally, we presented experimental results applying some of our techniques to a real-life dataset. This is the augmented journal version [23] of the conference paper that appeared in 2011.

6.26. On the genetic architecture of cytoplasmic incompatibility

Numerous insects carry intracellular bacteria manipulating their reproduction and thus facilitating their own spread. Cytoplasmic incompatibility (CI) is a common form of such manipulation, where a (currently uncharacterized) bacterial modification of male sperm 35 induces the early death of embryos unless the fertilized eggs carry the same bacteria, inherited from the mother. The death of uninfected embryos provides an indirect selective advantage to infected ones, thus enabling the spread of the bacteria. We used and expanded recently developed algorithms (the first being the one described in the previous item) to infer the genetic architecture underlying the complex incompatibility data from the mosquito *Culex pipiens*. We showed that CI requires more genetic determinants 40 than previously believed, and that quantitative variation in gene products potentially contributes to the observed CI patterns. In line with population genetic theory of CI, our analysis suggests that toxin factors (those inducing embryo death) are present in fewer copies in the bacterial genomes than antitoxin factors (those ensuring that infected embryos survive). In combination with comparative genomics, our approach will provide helpful guidance to 45 identify the genetic basis of CI, and more generally of other toxin / anti-toxin systems that can be conceptualised under the same framework. This work is currently submitted for publication. It was done in collaboration with Sylvain Charlat from the LBBE.

6.27. Viral population structure and dynamics

The work which started a few years ago with the Pasteur Institute in Cambodia (Dr. P. Buchy) and the CIRAD at Montpellier (Dr. R. Frutos) on viral population structure and dynamics has been continued in 2012, focusing on the H5N1 and Dengue viruses. The exploratory statistical approach based on MCoA (see the Bamboo annual report for 2011) was used to identify a novel H5N1 endemic sub-clade specific to Cambodia [27] and the work performed last year on Dengue serotype 1 has been extended in 2012 to serotypes 2 and 3 [11] thus providing a more precise view of the virus population dynamics over the last 12 years and demonstrating "synchronized" replacements most probably linked to climatic disasters like flood or drought.

6.28. Charge group partitioning in biomolecular simulation

Molecular simulation techniques are increasingly being used to study biomolecular systems at an atomic level. Such simulations rely on empirical force fields to represent the intermolecular interactions. There are many different force fields availableleach based on a different set of assumptions and thus requiring different parametrization procedures. Recently, efforts have been made to fully automate the assignment of force-field parameters, including atomic partial charges, for novel molecules. In this work, we focused on a problem arising in the automated parameterisation of molecules for use in combination with the gromos family of force fields: namely, the assignment of atoms to charge groups such that for every charge group the sum of the partial charges is ideally equal to its formal charge. In addition, charge groups are required to have size at most k. We showed NP-hardness and gave an exact algorithm capable of solving practical problem instances to provable optimality in a fraction of a second [32].

BANG Project-Team

6. New Results

6.1. Proliferation dynamics and its control

6.1.1. Cell division dynamics in structured cell populations

Participants: José Luís Avila Alonso [DISCO project-team, Inria Saclay IdF], Annabelle Ballesta, Frédérique Billy, Frédéric Bonnans [Commands project-team, Inria Saclay IdF], Catherine Bonnet [DISCO project-team, Inria Saclay IdF], Jean Clairambault, Luna Dimitrio, Marie Doumic-Jauffret, Xavier Dupuis [Commands project-team], Olivier Fercoq [MaxPlus project-team, Inria Saclay IdF], Stéphane Gaubert [MaxPlus project-team, Inria Saclay IdF], Germain Gillet [IBCP, Université Cl. Bernard Lyon 1], Philippe Gonzalo [IBCP, Université Cl. Bernard Lyon 1], Pierre Hirsch [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Thomas Lepoutre [now in DRACULA project-team, Inria Rhône-Alpes, Lyon], Jonathan Lopez [IBCP, Université Cl. Bernard Lyon 1], Pierre Magal [University Bordeaux II], Anna Marciniak-Czochra [Institute of Applied Mathematics, Universität Heidelberg], Jean-Pierre Marie [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Roberto Natalini [IAC-CNR, Università Sapienza, Rome], Silviu Niculescu [DISCO project-team, Inria Saclay IdF], Hitay Özbay [Bilkent University, Ankara, Turkey], Benoît Perthame, Ruoping Tang [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Vitaly Volpert [CNRS Lyon, UMR5208, Camille Jordan Institute, Lyon], Jorge Zubelli [IMPA, Rio de Janeiro].

1. *Transition kernels in a McKendrick model of the cell division cycle.* This theme has continued to be developed with identification of model parameters by FUCCI imaging in collaboration with G. van der Horst's team in Amsterdam and with F. Delaunay's team in Nice, within the C5Sys European network, coordinated by F. Lévi (Villejuif) [10], [11], [12], [39], [42], [43]. Main young researchers on this theme, F. Billy has concluded her 2-year Inria postdoc at Bang, leaving for an industrial company in November 2012, and O. Fercoq (team MaxPlus, Saclay) has defended his PhD thesis at École Polytechnique in September 2012, only to leave for a postdoc position dedicated to optimisation theory in Edinburgh.
2. *Modelling haematopoiesis with applications to AML.* This theme has been active through a collaboration with Inria teams Commands (F. Bonnans, X. Dupuis) and Disco (JL Avila, C. Bonnet), and J.-P. Marie's team at St Antoine Hospital leukaemic tumour bank, where A. Ballesta, Cancéropole IdF-Inria postdoc has been detached (ending in March 2013) to identify parameters of a model of acute myeloblastic leukaemia (AML) in patient fresh cell cultures with and without anticancer drugs. This work has led to several presentations, and publications are in preparation.
3. *Hybrid models* Systems combining PDEs and discrete representations in hybrid models, with applications to cancer growth and therapy, in particular for AML, are the object of study of the ANR program *Bimod*, coordinated by V. Volpert (Lyon), associating CNRS (V. Volpert, Lyon), Bordeaux II University (P. Magal) and the Bang project-team.
4. *Molecular model of the activity of the p53 protein.* This work, the object of Luna Dimitrio's PhD thesis [1], co-supervised by J. Clairambault and R. Natalini (Rome), has led to her PhD defence in September 2012 at UPMC, and to a first publication [18], that should be followed by others. After L. Dimitrio's leave for the pharmaceutic industry, a new PhD student, Ján Eliš, has taken over this theme in September 2012 in a new PhD thesis at UPMC, under the supervision of J. Clairambault and B. Perthame

6.1.2. Physiological and pharmacological control of cell proliferation

Participants: Annabelle Ballesta, Frédérique Billy, Jean Clairambault, Sandrine Dulong [INSERM Villejuif (U 776)], Olivier Fercoq [MaxPlus project-team], Stéphane Gaubert [MaxPlus project-team], Thomas Lepoutre [Dracula project-team], Francis Lévi [INSERM Villejuif (U 776)].

1. *Periodic (circadian) control of cell proliferation in a theoretical model of the McKendrick type.* This theme (cf. supra "transition kernels...") has been continued [39], [11], [12], [10], [42], [43]. Whereas transition kernels between cell cycle phases without control have been experimentally identified in cell cultures by FUCCI imaging [12], their circadian control remains elusive and has been modelled on the basis of gating by plain cosines representing the influence exerted on these transition kernels by circadian clocks. To go further, it would be necessary to have access by cell imaging to the activity of the best physiological candidates to such gating, namely the cyclin-Cdk complexes, together with the activities of the clock-controlled proteins Wee1 and p21, which thus far have remained unavailable to us through biological experimentation with imaging.
2. *Intracellular pharmacokinetic-pharmacodynamic (PK-PD) models for anticancer drugs.* This theme has continued to be developed with new publications for the drugs irinotecan [40], [44], 5-fluorouracil and oxaliplatin [43].

6.1.3. Optimisation of cancer chemotherapy

Participants: Annabelle Ballesta, Frédérique Billy, Frédéric Bonnans [Commands project-team], Jean Clairambault, Sandrine Dulong [INSERM Villejuif (U 776)], Xavier Dupuis [Commands project-team], Olivier Fercoq [MaxPlus project-team], Stéphane Gaubert [MaxPlus project-team], Thomas Lepoutre [Dracula project-team], Alexander Lorz, Francis Lévi [INSERM U 776, Villejuif], Michael Hochberg [ISEM, CNRS, Montpellier], Benoît Perthame.

Optimising cancer chemotherapy, in particular chronotherapy, is the final aim of the activities mentioned above. This theoretical activity has been continued, using the McKendrick paradigm in works involving the *C5Sys* network [12], [42], [43], with numerical optimisation algorithms for the toxicity constraint, and also in more general settings taking into account another major issue of anticancer treatment, namely resistance to drugs in cancer cells. To this latter aim, we have developed another type of models based on integro-differential equations, which are inspired from those used in ecology for Darwinian evolution. These are aimed at studying another major issue in cancer therapy: appearance of resistances to treatment in tumour cell populations. Indeed, these cell populations, because of their heterogeneity and genomic instability, present an ability to adapt and evolve (in the Darwinian sense) that is much higher than in healthy cell populations [10], [27], [39]. The time scales under investigation, much shorter than in ecology, are still much longer than in microbiology, and are those of clinical treatments.

From a molecular point of view, studying drug resistance leads to the study of ABC transporters, which is one of the tracks followed by A. Ballesta, following her PhD thesis, in collaboration with F. Lévi's INSERM team in Villejuif [40], [44].

Underway is also the use of methods of optimal control developed by the Commands project-team (F. Bonnans, X. Dupuis) to optimise therapies in the treatment of Acute Myeloblastic Leukaemia (AML, cf. supra "Modelling haematopoiesis with applications to AML").

6.1.4. Protein polymerisation and application to amyloid diseases (ANR grant TOPPAZ)

Participants: Annabelle Ballesta, Vincent Calvez [ENS Lyon], Marie Doumic-Jauffret, Pierre Gabriel, Hadjer Wafaâ Haffaf, Benoît Perthame, Stéphanie Prigent [BPCP, INRA Jouy-en-Josas], Human Rezaei [BPCP, INRA Jouy-en-Josas], Léon Matar Tine [SIMPAF project-team, Inria Lille Nord-Europe].

Published in PLoS One in collaboration with the biologists' team of H. Rezaei [29], a new and very complete PDE model for protein polymerisation has been designed. Following F. Charles's work, A. Ballesta has applied this model to Huntington's disease (PolyQ expansion) and compared it with its ODE counterpart, leading to a better understanding of the leading mechanisms responsible for PolyQ fibrillisation. New applications of this framework model are in progress with H.W. Haffaf and S. Prigent.

The eigenvalue problem playing a major role in the representation of Prion proliferation dynamics and, in a more general way, of many fragmentation-coalescence phenomena, the article [15] published in J. de Math. Pur. Appl. investigated the dependency of the principal eigenvector and eigenvalue upon its parameters. We exhibited possible nonmonotonic dependency on the parameters, conversely to what would have been conjectured on the basis of some simple cases.

6.1.5. Inverse problem in growth-fragmentation equations

Participants: Marie Doumic-Jauffret, Marc Hoffmann [ENSAE], Nathalie Krell [Univ. Rennes I], Patricia Reynaud [CNRS, Nice Univ.], Lydia Robert [UPMC], Vincent Rivoirard [Paris IX Univ.], Léon Matar Tine [SIMPAF project-team, Inria Lille Nord-Europe].

In collaboration with statisticians (M. Hoffman, Professor at Université de Marne-la-Vallée, V. Rivoirard, MC at Université d'Orsay, and P. Reynaud, CR CNRS at Université de Nice), in the article [19] published in SIAM Num. Anal., we explored a statistical viewpoint on the cell division problem. In contrast to a deterministic inverse problem approach, we take the perspective of statistical inference. By estimating statistically each term of the eigenvalue problem and by suitably inverting a certain linear operator, we are able to construct an estimator of the division rate that achieves the same optimal error bound as in related deterministic inverse problems. Our procedure relies on kernel methods with automatic bandwidth selection. It is inspired by model selection and recent results of Goldenschluger and Lepski.

An extension of this work, which consists of the statistical estimation of a branching process modelling the same growth and fragmentation dynamics, has been submitted in [49], in collaboration with N. Krell, M. Hoffmann and L. Robert.

In [20], published in J. Math. Biol. with L. Matar Tine, we generalised the inverse techniques proposed previously in [53], [57], in order to adapt them to general fragmentation kernels and growth speeds. The potential applications of this problem are numerous, ranging from polymerisation processes to the cell division cycle. An extension of this work is in progress with M. Escobedo and T. Bourgeron.

6.2. Tissue growth, regeneration and cell movements

6.2.1. Chemotaxis, self-organisation of cell communities (KPP-Fisher and Keller-Segel)

Participants: Nikolaos Bournaveas [Univ. Edinburgh], Axel Buguin [UPMC, Institut Curie], Vincent Calvez [ENS Lyon], François James [univ. Orléans], Alexander Lorz, Grégoire Nadin [UPMC], Benoît Perthame, Jonathan Saragosti [Institut Curie], Pascal Silberzan [Institut Curie], Min Tang [Shanghai Jiaotong University], Nicolas Vauchelet.

Chemotaxis denotes the ability of some cells to undergo a directed movement in response to an extracellular chemical substance. A mathematical description of chemotaxis is a major issue in order to understand collective movements of bacterial colonies. Numerous mathematical models, at various scales, have been proposed, allowing for a good description of cell aggregation under chemotaxis at the macroscopic level, the first of all being that of Keller-Segel (1971), that is now at the centre of an abundant international scientific literature.

At the cell scale, one uses kinetic equations. Numerical simulations have been performed and blow-up is also observed, which differs highly from pointwise blow-up in parabolic models. Representing them leads to various theoretical questions and amounts to define measure solutions [25], [24] or to develop an existence theory.

6.2.2. Single-cell-based and continuum models of avascular tumours

Participants: Ibrahim Cheddadi, Dirk Drasdo, Benoît Perthame, Min Tang [Shanghai Jiaotong University], Nicolas Vauchelet, Irène Vignon-Clémentel [REO project-team].

The recent biomechanical theory of cancer growth considers solid tumours as liquid-like materials comprising elastic components. In this fluid mechanical view, the expansion ability of a solid tumour into a host tissue is mainly driven by either diffusion of cells (emerging on the mesoscopic scale by coarse graining from the cell micro-motility) or by cell division depending either on the local cell density (contact inhibition), on mechanical stress in the tumour, or both. For the two by two degenerate parabolic/elliptic reaction-diffusion system that results from this modelling, we prove there are always travelling waves above a minimal speed and we analyse their shapes. They appear to be complex with composite shapes and discontinuities. Several small parameters allow for analytical solutions; in particular the incompressible cells limit is very singular and related to the Hele-Shaw equation. These singular travelling waves are recovered numerically. See [32].

6.2.3. Single cell-based models of tumour growth, tissue regeneration

Participants: Gregory Batt [CONTRAINTES project-team], François Bertaux, Géraldine Cellière, Chadha Chettaoui, Ibrahim Cheddadi, Dirk Drasdo, Adrian Friebel, Rolf Gebhardt [Univ. of Leipzig, Germany], Adriano Henney [Director Virtual Liver Network and VLN consortium], Jan G. Hengstler [Leibniz Research Centre, Dortmund, Germany and CANCERSYS consortium], Stefan Höhme, Elmar Heinzle [University of Saarbrücken and NOTOX consortium], Isabelle Hue [INRA], Nick Jagiella, Ursula Klingmüller [German Cancer Centre, Heidelberg and LungSys Consortium], Axel Krinner, Johannes Neitsch, Benoît Perthame, Ignacio Ramis-Conde, Luc Soler [IRCAD, Coordinator EU-project PASSPORT and PASSPORT consortium], Jens Timmer [University of Leipzig, Germany], Irène Vignon-Clémentel [REO project-team], Juhui Wang [INRA], William Weens.

6.2.3.1. A Multi-scale model for clonal competition in growing tumours

In this work we set up a multi-scale model testing the impact of three experimentally found variants of a signal transduction pathway controlling cell-cell adhesion on multi-cellular growth as well as the possible consequences of inhomogeneous populations where each of the three phenotypes competed [30].

6.2.3.2. Growth of cell populations in embedding granular and cell-like matter

In this work simulations of growing 2D and 3D clones embedded in granular and cell-like matter were mimicked [21]. The influence of active directed cell motion vs. passive pushing triggered by cell proliferation, as well as of various parameters of the embedding matter, such as the friction of embedding objects with its environment, adhesion strength, size of objects, elastic modulus etc. on the growth kinetics and the spatial pattern has been studied. The emerging patterns are strongly reminiscent of a fingering instability (a type of a Saffman-Taylor instability) occuring if a viscous fluid is injected into a more viscous fluid constrained between two plates (Hele Shaw cell).

6.2.3.3. Quantitative modelling of multi-cellular spheroids

Nick Jagiella in his thesis has worked out how stepwise and iteratively mechanisms controlling the spatial-temporal growth dynamics can be inferred by combining information from bright field micrographs stained for proliferating, dying cells, cell nuclei and extra-cellular matrix with the macroscopic growth kinetics.

This thesis, pursued within the German network project LUNGSYS was defended in September 2012. The thesis work was mainly supervised by Dirk Drasdo, PI for this part within the LUNGSYS project. Main collaborators were Margareta Mueller (previously DKFZ, Heidelberg) and Ursula Klingmueller, (DKFZ Heidelberg).

Moreover, Géraldine Cellière has worked out a model to mimic the aggregation of cells in the hanging drop method, a standard method to generate 3D multi-cellular aggregates. The kinetics and final configuration give information on multicellular aggregates. This work is pursued within the EU NOTOX project. Main collaborators are Fozia Noor and Elmar Heinzle (Univ. of Saarbruecken).

6.2.3.4. Image reconstruction of 3D liver architecture at subcellular level

In order to permit simulation liver function we started to set up an image processing pipeline resolving liver at subcellular scale. This will enable us to mimic all flows in liver, which comprises of blood flow through the micro-vessels (sinusoids), of blood plasma through the space between micro-vessel wall and hepatocytes, the main type of liver cells (called space of Disse), and of the bile through a network of bile canaliculi. Besides image analysis, also setting up the models of the flows has been started.

This work is conducted by the PhD student Adrian Friebel (IZBI, University of Leipzig) co-supervised by Dirk Drasdo and Stefan Hoehme (IZBI, University of Leipzig) within the Germany funded grant project Virtual Liver Network (VLN; PI from IZBI, Leipzig: Dirk Drasdo). Main collaborator is Jan G. Hengstler from the IfADo (directeur at the Leibniz Institute in Dortmund, Germany).

6.2.3.5. Ammonia metabolism during liver regeneration

Based upon the paper on liver regeneration after drug-induced damage (Hoehme et. al. PNAS 2010 [55]) we in a next step investigated the change of ammonia metabolism during the regeneration process. Ammonia is toxic for the body. We linked our spatial-temporal liver lobule model with a compartment model for the ammonia, glutamine and urea metabolism. In the latter we consider a compartment (the peri-central compartment) in which glutamine synthetase, a strongly ammonia-detoxifying enzyme, is degraded efficiently and a (peri-portal) compartment, in which this is not the case. By testing different hypotheses on the chemical reactions taking place during the degradation process and quantitatively comparing to time-space data of the regeneration process including data on the activity of glutamine synthetase we were able to propose a potentially missing chemical reaction. Validation experiments have been started and suggest that the original reaction scheme was indeed incomplete.

This work is conducted by Dirk Drasdo and Stefan Hoehme (IZBI, University of Leipzig) partly within the Germany funded grant project Virtual Liver Network (VLN; PI from IZBI, Leipzig: Dirk Drasdo) and the EU project NOTOX. Main collaborators are Rolf Gebhardt (chair for Biochemistry, University of Leipzig), Jan G. Hengstler from the IfADo (Leibniz Institute in Dortmund, Germany) and BioControl Jena GmbH, a company in Jena, Germany.

6.2.3.6. Multi-scale simulation of cell cycle progression during liver regeneration

In previous work on liver regeneration after drug induced damage (Hoehme et. al. PNAS 2010 [55]) the experimentally observed spatial-temporal proliferation pattern has been used as an input parameter. We have now started to study the molecular control of cell cycle progression by hepatocyte growth factor (HGF). Based on model predictions with a hypothesized model linking the downstream activation of the HGF-pathway with cell cycle progression, experiments were performed which now led to a validated intracellular model of cell cycle progression by HGF. Moreover, based on model simulations predicting that two sources of HGF are necessary to explain the experimentally observed proliferation pattern, experiments detecting the potential sources of HGF have been initiated. The models are multi-scale in that the precise spatial architecture of a piece of liver tissue is modelled representing each individual hepatocyte as well as the blood micro-vessels. A system of ODE's mimicking the HGF signalling and its impact on cell cycle progression is solved inside each individual cell. The project works out a systematic strategy to stepwise identify multi-scale multi-level processes in tissue organisation extending the lines pursued in Hoehme et. al. [55] and Holzhuetter et. al. [23].

This work is conducted by Dirk Drasdo and Stefan Hoehme (IZBI, University of Leipzig) within the Germany funded grant project Virtual Liver Network (VLN; PI from IZBI, Leipzig: Dirk Drasdo). Main collaborators are Ursula Klingmueller and Lorenza D'Alessandro (UK is Professor at Heidelberg University and department head at German Cancer Research Centre (DKFZ), Heidelberg, Germany) as well as Jens Timmer and Andreas Raue (JT is Professor University of Freiburg, Germany).

6.2.3.7. Phenotypes in early liver cancer

The model of a liver lobule, the smallest functional unit of liver (Hoehme et. al., PNAS 2010 [55]) has been used as a starting point to explain the experimentally observed early tumour phenotypes. We made a sensitivity analysis to identify the parameters that influence the tumour phenotype. Each simulation mimicked a monoclonal tumour. We could show that the observed early phenotypes could be explained by only a few sensitive parameters which are the direction of cell division, cell-micro-vessel adhesion, and destruction of micro-vessels by the tumour cells.

This work has been taken over from the previous PhD student William Weens by the PhD student François Bertaux who is co-supervised by Dirk Drasdo and Gregory Batt. Main collaborator is Jan G. Hengstler from the IfADo (directeur at the Leibniz Institute in Dortmund, Germany).

6.2.3.8. Regeneration of liver after partial hepatectomy

We continued this earlier activity by initiating experiments on pigs to test the model prediction that the 2nd wave of proliferation during regeneration after partial hepatectomy in pig should occur only close to the Glisson capsule, that encloses the liver, while in mouse proliferation occurs homogeneously and isotropically distributed over the whole liver lobe.

This work is conducted by Dirk Drasdo and Stefan Hoehme (IZBI, University of Leipzig) within the Germany funded grant project Virtual Liver Network. Main collaborators are Jan G. Hengstler from the IfADo (Leibniz Institute in Dortmund, Germany) and Eric Vilbert, Centre Hépato-Biliaire (CHB)- INSERM U785, Hospital Paul Brousse, Villejuif.

6.2.3.9. High resolution model for eukaryotic cells

In order to permit simulations directly out of 3D reconstructions of confocal laser scanning micrographs at subcellular resolution we developed a model that is capable to resolve complex cell shapes. The model parameters were calibrated by comparison with experiments probing the material properties of cells. Moreover, the cell division was implemented. The model was integrated into the CellSys software.

This work is conducted by the PhD student Johannes Neitsch (IZBI, University of Leipzig) co-supervised by Dirk Drasdo and Stefan Hoehme (IZBI, University of Leipzig) within the Germany funded grant project Virtual Liver Network (VLN). Main collaborators are Jan G. Hengstler from the IfADo (directeur at the Leibniz Institute in Dortmund, Germany) and Josef Kaes (Prof. for Experimental Physics, Univ. Leipzig).

6.2.3.10. Yeast cells playing the Game of Life

Within a collaboration with a synthetic biology lab at MIT, we work on the multicellular modelling of engineered yeast cell populations. Those cells secrete a messenger molecule (IP) which diffuse in the medium, bind to other cells, and trigger a signalling cascade which finally induce expression of lethal genes. A model has been established based on our single-cell-based model framework associated with PDE's simulations, and it is currently used to explain and guide experiments obtained at MIT.

This work is conducted within the project Sine2Arti by François Bertaux co-supervised by Gregory Batt and Dirk Drasdo, and by Szymon Stoma. Main collaborator is Ron Weiss, MIT, Boston, USA.

6.2.3.11. Stochastic modelling of extrinsic apoptosis

Here we extended a well-established ODE model of TRAIL-induced apoptosis developed by Sorger's group in Harvard by the possible effect of cell-to-cell variability due to stochasticity of rare events in the cascade.

This work is conducted within the project Sine2Arti by François Bertaux co-supervised by Gregory Batt and Dirk Drasdo, and by Szymon Stoma as well as Xavier Duportet for the experimental part.

6.2.3.12. Artificial Homeostasis in HeLa cells

The aim is to genetically engineer human cancer cells (HeLa cell line) such that they perform population control in a petri dish. To do so, it is made use of extrinsic apoptosis by forcing cells to produce a messenger molecule able to trigger apoptosis above a certain threshold concentrations in the medium. We developed a mathematical model which integrates both PDEs and intracellular components into a single-cell-based model framework. Such model allows to help designing the genetic system that should be integrated into cells as well as guiding experiments.

This work is conducted within the project Sine2Arti by François Bertaux who is co-supervised by Gregory Batt and Dirk Drasdo. Moreover Szymon Stoma for the modelling part, as well as Xavier Duportet for the experimental part from the CONTRAINTES team are included.

6.2.4. Modelling flow in tissues

Participants: Lutz Brusch [TU Dresden], Dirk Drasdo, Adrian Friebel [IZBI, University of Leipzig], Stefan Hoehme [IZBI, University of Leipzig], Nick Jagiella [Inria and IZBI, University of Leipzig], Hans-Ulrich Kauczor [University of Heidelberg, Germany], Fabian Kiessling [University Clinics, Technical University of Aachen, Germany], Ursula Klingmueller [German Cancer Research Centre (DKFZ), Heidelberg, Germany], Hendrik Laue [Fraunhofer Mevis, Bremen, Germany], Ivo Sbazarini [MPI for Molecular Cell Biology and Genetics, Dresden, Germany], Irène Vignon-Clémentel [REO project-team], Marino Zerial [MPI for Molecular Cell Biology and Genetics, Dresden, Germany].

6.2.4.1. Flow and perfusion scenarios in cancer

In this subject we simulated typical flow and perfusion scenarios in tumour and tissue including, how the spatial-temporal pattern look like on the scale of non-invasive medical image modalities currently applied, to infer parameters that are used to or may permit to evaluate the perfusion of tumors in patients. The simulations use Poiseuille flow and Kirchhoff rule in 3D blood network representing typical architectures.

The work was part of the PhD thesis of Nick Jagiella, defended in September 2012 co-supervised by Dirk Drasdo and Irene Vignon-Clementel, and conducted within the grant funded network projects LUNGSYS and LUNGSYS II. Main collaborators were Oliver Sedlaczek, DKFZ Heidelberg and University of Heidelberg, Fabian Kissling, Technical University of Aachen and Hendrik Laue, Fraunhofer Mevis, Bremen (all in Germany).

6.2.4.2. Flow in liver lobules

The aim of this project is to simulate realistically the flow of matter within liver lobules from images generated with different image modalities at histological scales. So far we have established a model of blood flow and perfusion in liver lobules based upon 3D reconstruction of confocal micrographs.

This work is conducted by collaboration of different groups within the Germany funded grant project Virtual Liver Network. From our group Nick Jagiella, Adrian Friebel, and Stefan Hoehme, Dirk Drasdo are involved, main collaborators are Irene Vignon-Clementel (REO project team Inria), Marino Zerial and Ivo Sbazarini (Max-Planck Institute for Molecular Cell Biology and Genetics, Dresden, Germany), Lutz Brusch (Technical University of Dresden) and Jan G. Hengstler from the IfADo (Leibniz Institute in Dortmund, Germany).

6.2.5. Contraction of acto-myosin structures in morphogenesis and tissue repair

Participants: Luís Almeida, P. Bagnerini [Univ. Genova], A. Habbal [Univ. Nice], A. Jacinto [CEDOC, Lisbon], M. Novaga [Univ. Padova], A. Chambolle [École Polytechnique], J. Demongeot [Univ. Grenoble].

Contraction of actin structures (in one, two or three dimensions) plays an important role in many cellular and tissue movements, both at a multicellular tissue level and at a cellular (and even intracellular) one: from muscle contraction to neural tube closure, epiboly in zebrafish embryo, the contractile ring in cytokinesis, cell crawling,... examples are everywhere in the living world. These structures consist of meshworks of actin filaments (which are like fibers) that are cross-linked by molecular motors (Myosin II) which can make the actin filaments slide relative to each other, thus generating deformation movements.

In [4] we are particularly interested in modelling the contraction of acto-myosin cables in morphogenesis and tissue repair. The experiments done in collaboration with A. Jacinto's lab show that the local curvature (and in particular its sign) plays an important role in the contractile behaviour of the acto-myosin cables. These experimental results led us to develop some of these ideas in [6] and to do a more abstract study of flows by the positive part of the curvature in [5].

6.3. Neurosciences

Participants: M. Galtier, G. Hermann, M. Magnasco, T. Taillefumier, Jonathan Touboul.

We pursued the analysis of the dynamics of networks of neurons in the presence of noise. Limit theorems in simple cases were treated in [9], and more refined models including space, delays and heterogeneities were analysed in [34], [35], toubouldelays:12,touboulNeuralFieldsDynamics:12. In all these contributions we analysed the eminently important role of noise and heterogeneity on the qualitative dynamics of networks. Mathematical results were obtained for representation of the solutions to linear functional differential equations [22] that were motivated by plasticity phenomena in the cortex.

6.4. Free surface geophysical flows

Participants: Emmanuel Audusse [LAGA - Université Paris 13, Institut Galilée], Anne-Céline Boulanger, Marie-Odile Bristeau, Benoît Perthame, Jacques Sainte-Marie, Nicolas Seguin, Edwige Godlewski, Anne Mangeney, Yohan Penel, Raouf Hamouda, Philippe Ung.

The ANGE team has been created in november 2012. This new team (led by J. Sainte-Marie) resumes the activities of the BANG team concerning geophysical flows.

We are involved in research concerning the numerical simulation of free surface geophysical flows such as rivers, lakes, coastal areas and also overland flows. Many applications related to environmental problems are concerned : floodings, dam breaks, swell, transport and diffusion of pollutants, water quality, upwellings, sustainability of aquatic ecosystems, ...

The basic model for these problems is the 3D free surface Stokes system leading to a 3D solver [52] with a moving mesh. However for efficiency reasons, vertically averaged models such as the Saint-Venant system [54] are often used.

The Saint-Venant equations are deduced of the Navier-Stokes system with two main assumptions:

- the pressure is hydrostatic,
- the horizontal velocity is represented by its average.

We have developed extensions of the Saint-Venant system where the basic Saint-Venant solver [51] is still used and, in that way, the robustness, the efficiency and the easiness to treat the free surface are preserved while the domain of validity is larger.

In these extensions, we relax the two above assumptions. Actually, we have derived a non-hydrostatic shallow water model and a multilayer Saint-Venant system.

We have coupled the hydrodynamics of free surface flows with other phenomena such as biology (phytoplankton culture) or erosion.

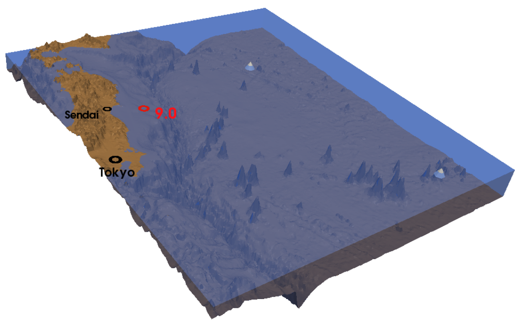


Figure 1. Map of Japan with the seism epicentre and the DART buoys 21418 and 21413.

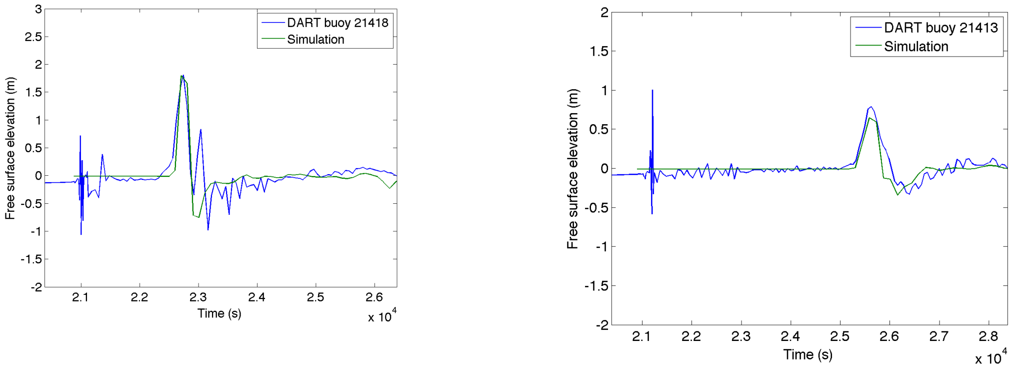


Figure 2. Free surface elevation of the sea, comparison between the recorded data by the buoys 21418 and 21413 and the simulation obtained with our 3d Navier-Stokes code.

BEAGLE Team

6. New Results

6.1. Model of genome reduction

To test whether the effect of the rearrangement rate on genome size holds independently of the artificial chemistry of the aevol (individual-based) model, we have written a simpler, mathematical model of genome size evolution including both genes and intergenic sequences, evolving through small insertions and deletions, large deletions and duplications and through selection based on gene number. The approach was presented this summer as a poster at the SMBE conference (Society for Molecular Biology and Evolution). We have shown analytically that without selective pressure, genomes spontaneously shrink and that large genomes are particularly unstable. When selection is included that favors the highest gene number, simulations show that genome sizes do not grow indefinitely as large genomes cannot be sustained. There is a trade-off between fitness and structural stability. A manuscript is being written and will be submitted in January.

6.2. The Paradoxical Effects of Allelic Recombination on Fitness

D.P. Parsons, C. Knibbe, G. Beslon. [42]

We introduced in the aevol model the possibility of DNA exchange by allelic recombination, in order to study the influence of recombination on the evolution of both fitness and genomic architecture. Surprisingly, despite the theoretical benefits it could confer, there seems to be very little (if any) differences in the fitness of the evolved organisms between the different groups of simulations.

6.3. Genome histories reconstructions

E. Tannier (Beagle), with B. Boussau, G. Szollosi, V. Daubin, L. Duret, M. Gouy, S. Abby (LBBE, Lyon), N. Lartillot (Univ Montreal), C. Chauve (SFU Vancouver)

Lateral gene transfer has been discovered in the 1940's and since has been seen by phylogeneticists as a noise one had to remove before analyses in molecular evolution. Thi noise was recently considered so important that it would blur the historical signal and leave no hope for reconstructing a phylogeny. In a series of papers [16], [31], [32], we model the lateral gene transfer and prove that it can be used as a signal to

- reinforce the support for the phylogeny of vertical descent [16]
- order in time some bacterial diversification events, and thus provide a unique source for dating the history of life (more than 3/4 of it is prokaryotic and the fossil record is not abundant) [31]
- have a trace of extinct species which did not leave any descendants, if they gave some genes to more successful lineages, which opens the way to include them in molecular phylogenies [32]

We devised methods to trace whole genome evolution, with multi-scale mutations: from nucleotide substitutions to large-scale rearrangements. We provided a mammalian phylogeny accounting fot the evolution of several thousand genes [17], and a method to sample among evolutionary scenarios [27].

Eventually we built a model of evolution of relations between pairs of genes, enable us to reconstruct ancestral genome structure or ancestral systems of interactions [18]. In the case of genome structure we also published a method to linearize a set of ancestral relations [26].

6.4. A Theory of Rate Coding Control by Intrinsic Plasticity Effects

H. Berry (Beagle), J. Naudé and B. Delord (ISIR, CNRS UMR 7222, Univ P&M Curie, Paris) and J.T. Paz (Stanford Univ Medical Center, CA, USA).

Over the past decades, experimental and theoretical studies of the cellular basis of learning and memory have mainly focused on synaptic plasticity, the experience-dependent modification of synapses. However, behavioral learning has also been correlated with experience-dependent changes of non-synaptic voltage-dependent ion channels. This intrinsic plasticity changes the neuron's propensity to fire action potentials in response to synaptic inputs. Thus a fundamental problem is to relate changes of the neuron input-output function with voltage-gated conductance modifications. Using a sensitivity analysis in biophysically realistic models, we depicted a generic dichotomy between two classes of voltage-dependent ion channels [28]. These two classes modify the threshold and the slope of the neuron input-output relation, allowing neurons to regulate the range of inputs they respond to and the gain of that response, respectively. We further provide analytical descriptions that enlighten the dynamical mechanisms underlying these effects and propose a concise and realistic framework for assessing the computational impact of intrinsic plasticity in neuron network models. Our results account for a large repertoire of empirical observations and may enlighten functional changes that characterize development, aging and several neural diseases, which also involve changes in voltage-dependent ion channels.

6.5. The influence of topology on calcium wave propagation in 3D astrocyte networks

H. Berry, Jules Lallouette (Beagle)

Glial cells are non-neuronal cells that constitute the majority of cells in the human brain and significantly modulate information processing via permanent cross-talk with the neurons. Astrocytes are also themselves inter-connected as networks and communicate via chemical wave propagation. How astrocyte wave propagation depends on the local properties of the astrocyte networks is however unknown. We have investigated the influence of the characteristics of the network topology on wave propagation [38]. Using a model of realistic astrocyte networks (> 1000 cells embedded in a 3D space), we show that the major classes of propagations reported experimentally can be emulated by a mere variation of the topology. Our study indicates that calcium wave propagation is favored when astrocyte connections are limited by the distance between the cells, which means that propagation is better when the mean-shortest path of the network is larger. This unusual property sheds new light on consistent reports that astrocytes in vivo tend to restrict their connections to their nearest neighbors.

6.6. Dynamics of protein aggregation in Escherichia coli

H. Berry, Anne-Sophie Coquel (Beagle) and Ariel Lindner (INSERM U1001, Cochin Medical School, Paris).

Protein aggregation plays a key role in cell decline and leads to several human disease linked to ageing like Alzheimer or Parkinson disease and prion disease. In Escherichia coli bacteria, accumulation of damaged proteins and their asymmetric segregation allowed to show ageing signs. This work [14] is focused on the in vivo spatial dynamics of protein aggregates in E. coli. Protein aggregates can be classified as inclusion bodies and they are amorphous or amyloid with a high order level due to β sheets. Combining a double theoretical and experimental approach, based on modeling and time-lapse and microfluidic microscopy, we studied the mechanism governing the motion of protein aggregates and the long-term vertical transmission of prionoid aggregates for about 10 generations. Our results show clearly that Brownian diffusion governs the motion of protein aggregates and the diffusion coefficient depends on the molecule size. The amyloid proteinopathy study shows the existence of lineages propagating two kind of aggregates : globular or comet-like. Lineages maintaining globular aggregates present an increase of the aggregate size until inhibition of the growth rate while comet-like aggregates are mildly detrimental to growth. We observed also at low frequency in some lineages the presence of both aggregates and a switch between them. Globular foci give born to comet-like aggregates.

6.7. Model of membrane domains emergence

HA Soula, A Coulon, G Beslon (Beagle)

In the classical view, cell membrane proteins undergo isotropic random motion, that is a 2D Brownian diffusion that should result in an homogeneous distribution of concentration. It is, however, far from the reality: Membrane proteins can assemble into so-called microdomains (sometimes called lipid rafts) which also display a specific lipid composition. The amount of this so-called overconcentration at equilibrium is simply related to the ratio of diffusion coefficients between zones of high and low diffusion. Expanding the model to include particle interaction, we show that inhomogeneous diffusion can impact particles clusterization as well. The clusters of particles were more numerous and appear for a lower value of interaction strength in the zones of low diffusion compared to zones of high diffusion. Provided we assume stable viscosity heterogeneity in the membrane, our model proposes a simple mechanism to explain particle concentration heterogeneity and hence domains.

6.8. Deleterious effect of receptor clustering on canonical signaling pathways

BR Caré, HA Soula (Beagle)

Classical framework for analyzing system biology pathways assumed that the cells are a well mixed and stirred medium. This hypothesis can dramatically fail in the case of membrane based stage of signaling. Due to microdomains membrane receptors are colocalized. Using individual based-model we show that this clustering seriously impairs the overall ligands binding as well as several pathways downstream. We contend that this unexpected effect is a very simple tool available for a cell to adjust its response.

6.9. Novel mathematical model of Adipose tissue cells size distribution

HA Soula (Beagle) C. Soulage, A Géloën (CARMEN)

We present a novel model to explain bimodality of size distribution of adipocytes: adipose tissue cells. These cells are dedicated to storing energy excess in form of fat and therefore can experience wide variations of sizes. Ubiquitous to all the species, we tested so far the size distributions are bimodal with no characteristic size. Using data from experiments, we provide a simple surface based model of circulating fats that cells can exchange. We show that in the physiological range for the parameters of the model, we obtain bimodal distribution. We also provide prediction of the size evolution during severe caloric restriction that we were able to verify experimentally as well.

BIGS Project-Team

6. New Results

6.1. Modern methods of data analysis

Participants: R. Bar, B. Lalloué, J-M. Monnez, C. Padilla, D. Zmirou, S. Deguen.

In 2012, our contributions to data analysis in a Biological context are twofold:

- At a theoretical level, we have kept on working on the so-called online data analysis alluded to at the *Scientific Foundations* Section. Specifically we have carried on in [15] (see also [4]) the analysis of data whose characteristics such as mathematical expectation or covariance matrix may vary with time, a problem which arises very naturally in this context. Moreover, in order to save computation time and thus take into account more data, a method considering several data at each step (we talk about data blocks) is proposed. This technique can also be useful if data are sent and received block-wise. In parallel, a R package performing most of the methods of factorial analysis in an online way is under development.
- At a practical level, our efforts have focused (cf. [19]) on an interesting study concerning the construction of a socio-economic neighborhood index which might quantify health inequalities. While several socio-economic indices already exist in this application field, most of them are very simple both in term of methodological construction and of number of variables taken into account, and only a few use data mining techniques. In order to exploit the large data sets of socio-economic variables provided by censuses and create neighborhood socio-economic indices yielding a better highlight of social health inequalities, a procedure was set in order to automatically select the best indicators in a set of socio-economic variables and synthesize them in a quantitative index. Application to three French metropolitan areas allowed testing the procedure and confirming both its reproducibility on various urban areas and the quality of the neighborhood socio-economic indices we had created (according to field experts and study partners). In this context, our expertise in data analysis allows for a good prediction by means of rigorous methods. Eventually, in order to simplify the application of the creation procedure of a socio-economic index for non-statisticians, a R package called SesIndexCreatoR was created to implement it.
- Publication of the sharp results obtained in [8] on local regression techniques.

6.2. Tumor growth modeling

Participants: R. Keinj, T. Bastogne, P. Vallois.

Up to now, the treatment planning systems used in radiotherapy only use mathematical models to describe the delivery of physical doses of radiation within biological tissues but cannot accurately predict the biological damages caused by such treatments. One important bottleneck is to account for the cell damage heterogeneity in the treated tumor. To this aim we firstly introduced in [51] a stochastic model based on multi-state Markov chains able to describe both treatment damage and cell reparation process.

More recently, we have proposed another model describing the lifespan of heterogenous tumors treated by radiotherapy. It is a bi-scale model in which the cell and tumor lifespans are represented by random variables. First and second-order moments, as well as the cumulative distribution functions and confidence intervals are expressed for the two lifespans with respect to the model parameters. One interesting result is that the mean value of the tumor lifespan can be approached by a logarithmic function of the initial cancer cell number. Moreover, we show that TCP (Tumor Control Probability) and NTCP (Normal Tissue Complication Probability), used in radiotherapy to evaluate, optimize and compare treatment plans, can be derived from the tumor lifespan and the surrounding healthy tissue respectively. Finally, we propose a ROC curve, entitled ECT (Efficiency-Complication Trade-off), suited to the selection by clinicians of the appropriate treatment planning (see [10]).

One difference between photodynamic therapy (PDT) and radiotherapy (RT) is the irradiation signal (X ray in RT and light beam in PDT). Another one is the treatment planning: 10 to 30 daily sessions of treatment in RT against only one for PDT. To adapt the previous model to PDT, a continuous-time version was developed and proposed in [18]. The model has been implemented into Matlab and numerical simulations have emphasized the effects of the model parameters on the model output.

In the framework of a new collaboration with S. Niclou (NorLux Neuro-Oncology Laboratory, Department of Oncology, Centre de Recherche Public de la Santé, Luxembourg), we have extended our stochastic model of cell damage to describe the phenotypic heterogeneity in brain tumors. Preliminary results have recently been presented in [16]. Cancer stem cell (CSC) hypothesis suggests that tumor progression and recurrence rely on a small subpopulation of cancer cells with stem-like properties. The unresolved question is whether cancer stem cells lead to organisation of intratumoral phenotypic heterogeneity by hierarchical differentiation events or whether they represent one of the transitory phenotypic states. This is crucial not only for our understanding of tumor progression, but also for the successful design of novel therapeutic strategies targeting CSCs. Let us also highlight the fact that those studies are related to a more application oriented research synthesized in [3], [13], [21]

6.3. Piecewise deterministic Markov processes

Participants: A. Crudu, A. Debussche, A. Muller-Gueudin, O. Radulescu.

Piecewise deterministic Markov processes are models which feature in a prominent way in Biomedical applications. They appear in two contributions of our team this year.

(1)*Convergence of stochastic gene networks.* In [24], [5], we propose simplified models for the stochastic dynamics of gene network models arising in molecular biology. Those gene networks are classically modeled by Markov jump processes, which are extremely time consuming. To overcome this drawback, we study the asymptotic behavior of multiscale stochastic gene networks using weak limits of Markov jump processes.

We consider a set of chemical reactions R_r, $r \in \mathcal{R}$; $\mathcal{R}$ is supposed to be finite. These reactions involve species indexed by a set $S = 1, \cdots, M$, the number of molecules of the species i is denoted by n_i and $X \in \mathbb{N}^M$ is the vector consisting of the n_i's. Each reaction R_r has a rate $\lambda_r(X)$ which depends on the state of the system, described by X and corresponds to a change $X \to X + \gamma_r$, $\gamma_r \in \mathbb{Z}^M$.

Mathematically, this evolution can be described by the following Markov jump process. It is based on a sequence $(\tau_k)_{k\geq 1}$ of random waiting times with exponential distribution. Setting $T_0 = 0$, $T_i = \tau_1 + \cdots + \tau_i$, X is constant on $[T_{i-1}, T_i)$ and has a jump at T_i. The parameter of τ_i is given by $\sum_{r\in\mathcal{R}} \lambda_r(X(T_{i-1}))$:

$$\mathbf{P}(\tau_i > t) = \exp\left(-\sum_{r\in\mathcal{R}} \lambda_r(X(T_{i-1}))t\right).$$

At time T_i, a reaction $r \in \mathcal{R}$ is chosen with probability $\lambda_r(X(T_{i-1}))/\sum_{r\in\mathcal{R}} \lambda_r(X(T_{i-1}))$ and the state changes according to $X \to X + \gamma_r$: $X(T_i) = X(T_{i-1}) + \gamma_r$. This Markov process has the following generator:

$$Af(X) = \sum_{r\in\mathcal{R}} [f(X+\gamma_r) - f(X)]\, \lambda_r(X).$$

In the applications we have in mind, the numbers of molecules have different scales. Some of the molecules are in small numbers and some are in large numbers. Accordingly, we split the set of species into two sets C and D with cardinals M_C and M_D. This induces the decomposition $X = (X_C, X_D)$, $\gamma_r = (\gamma_r^C, \gamma_r^D)$. For $i \in D$, n_i is of order 1 while for $i \in C$, n_i is proportional to N where N is a large number. For $i \in C$, setting $\widetilde{n}_i = n_i/N$, $\widetilde{n}_i$ is of order 1. We define $x_C = X_C/N$ and $x = (x_C, X_D)$.

For this kind of system, we are able to give in [5] some relevant information on the asymptotic regime $N \to \infty$ when different type of reactions are involved. Depending on the time and concentration scales of the system we distinguish four types of limits:

- Continuous piecewise deterministic processes (PDP) with switching.
- PDP with jumps in the continuous variables.
- Averaged PDP.
- PDP with singular switching.

We justify rigorously the convergence for the four types of limits.

(2)*Variable length Markov chains.* A classical random walk $(S_n,\, n \in \mathbb{N})$ is defined by $S_n := \sum_{k=0}^{n} X_k$, where (X_k) are i.i.d. When the increments $(X_k)_{k\in\mathbb{N}}$ are a one-order Markov chain, a short memory is introduced in the dynamics of (S_n). This so-called "persistent" random walk is no longer Markovian and, under suitable conditions, the rescaled process converges towards the integrated telegraph noise (ITN) as the time-scale and space-scale parameters tend to zero (see [70], [71], [50]). The ITN process is effectively non-Markovian too. In [28] our aim has been to consider persistent random walks (S_t) whose increments are Markov chains with variable order which can be infinite.

Associated with a process (X_n) which takes its values in a finite set, we consider an integer valued process (M_n) so that (X_n, M_n) is Markov and M_n measures the size of the memory at time n. This variable memory is justified by a one-to-one correspondence between (X_n) and a suitable Variable Length Markov Chain (VLMC), since for a VLMC the dependency from the past can be unbounded. We prove in [28] that, under a suitable rescaling, (S_n, X_n, M_n) converges in distribution towards a time continuous process $(S^0(t), X(t), M(t))$. The process $(S^0(t))$ is a semi-Markov and Piecewise Deterministic Markov Process whose paths are piecewise linear.

Observe that, though our study in [28] is made at a theoretical level, it leads to potentially interesting applications in growth models for tumors. This kind of link will be developed in the next future.

6.4. Inference for Gaussian systems

Participants: T. Cass, S. Cohen, M. Hairer, C. Litterer, F. Panloup, L. Quer, S. Tindel.

As mentioned at the *Scientific Foundations* Section, the problem of estimating the coefficients of a general differential equation driven by a Gaussian process is still largely unsolved. To be more specific, the most general ($\mathbb{R}$-valued) equation handled up to now as far as parameter estimation is concerned (see [69]) is of the form:

$$X_t^\theta = a + \theta \int_0^t b(X_u)\, du + B_t,$$

where θ is the unknown parameter, b is a smooth enough coefficient and B is a one-dimensional fractional Brownian motion. In contrast with this simple situation, our applications of interest (see the *Application Domains* Section) require the analysis of the following $\mathbb{R}^n$-valued equation:

$$X_t^\theta = a + \int_0^t b(\theta; X_u)\, du + \int_0^t \sigma(\theta; X_u)\, dB_t, \tag{29}$$

where θ enters non linearly in the coefficient, where σ is a non-trivial diffusion term and B is a d-dimensional fractional Brownian motion. We have thus decided to tackle this important scientific challenge first.

To this aim, here are the steps we have focused on in 2012:

- An implementable numerical scheme for equations driven by irregular processes, which is one of the ingredients one needs in order to perform an accurate statistical estimation procedure (see [6]).
- A better understanding of the law of the solution X_t^θ to equation (1), carried out in [25]. This step allows to obtain smoothness of density for our equation of interest in a wide range of contexts, which is an essential prerequisite for a good estimation procedure.
- Another important preliminary step for likelihood estimates for stochastic equations is a good knowledge of their invariant measure in the ergodic case. This is the object of our article [27].
- Finally we have also progressed in our knowledge of noisy differential systems by extending the range of applications of rough paths methods [11], [14].

BIOCORE Project-Team

6. New Results

6.1. Mathematical methods and methodological approach to biology

6.1.1. Mathematical analysis of biological models

Participants: Jean-Luc Gouzé, Olivier Bernard, Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, Francis Mairet, Rafael Muñoz-Tamayo, Elsa Rousseau.

6.1.1.1. Mathematical study of semi-discrete models

Semi-discrete models have shown their relevance in the modeling of biological phenomena whose nature presents abrupt changes over the course of their evolution [95]. We used such models and analysed their properties in several situations that are developed in 6.2.3 , most of them requiring such a modeling in order to take seasonality into account. Such is the case when the year is divided into a cropping season and a 'winter' season, where the crop is absent, as in our analysis of the sustainable management of crop resistance to pathogens [59] or in the co-existence analysis of epidemiological strains [19], [50]. Seasonality also plays a big role in the semi-discrete modeling required for the analysis of consumers' adaptive behavior in seasonal consumer-resource dynamics, where only dormant offspring survives the 'winter' [52].

6.1.1.2. Mathematical study of models of competing species

When several species are in competition for a single substrate in a chemostat, and when the growth rates of the different species only depend on the substrate, it is known that the generic equilibrium state for a given dilution rate consists in the survival of only one of the species. In [30], we propose a model of competition of n species in a chemostat, where we add constant inputs of some species. We achieve a thorough study of all the situations that can arise when having an arbitrary number of species in the chemostat inputs; this always results in a Globally Asymptotically Stable equilibrium where all input species are present with at most one of the other species.

The competition of several microalgal species was also studied in order to determine conditions that may give a competitive advantage to a species of interest. We study the competition for two species subject to photoinhibition at high light. This leads to a closed loop control strategy based on the regulation of the light intensity at the bottom of the reactor. The winning species is the one with the highest growth rate at high light. Then we show that the proposed controller allows the selection of a species of interest among n species [102].

6.1.2. Model design, identification and validation

Participants: Olivier Bernard, Francis Mairet.

One of the main families of biological systems that we have studied involves mass transfer between compartments, whether these compartments are microorganisms or chemical species in a bioreactor, or species populations in an ecosystem. We have developed methods to estimate the models of such systems [2]. These systems can be represented by models having the general structure popularized by [69], [74], and based on an underlying reaction network:

$$\frac{d\xi}{dt} = K\,r(\xi,\psi) + D(\xi_{in} - \xi) - Q(\xi)$$

We address two problems: the determination of the pseudo-stoichiometric matrix K and the modeling of the reaction rates $r(\xi,\psi)$.

In order to identify K, a two-step procedure has been proposed. The first step is the identification of the minimum number of reactions to be taken into account to explain a set of data. If additional information on the process structure is available, we showed how to apply the second step: the estimation of the pseudo-stoichiometric coefficients.

This approach has been applied to various bioproduction processes, among which activated sludge processes [68], anaerobic digestion [87], [114] and anaerobic digestion of microalgae [20]. Recently it was also used to reduce the ADM1 model in the case of winery effluent wastewater [88].

6.1.3. Nonlinear observers

Participants: Jean-Luc Gouzé, Olivier Bernard, Francis Mairet.

Interval observers

Interval observers give an interval estimation of the state variables, provided that intervals for the unknown quantities (initial conditions, parameters, inputs) are known [7]. We have extended the interval observer design to new classes of systems. First, we designed interval observers, even when it was not possible in the original basis, by introducing a linear, time-varying change of coordinates [105]. This approach was then extended to n-dimensional linear systems, leading to the design of interval observers in high dimensions [106]. Interval observers for non linear triangular systems satisfying Input to State Stability has been proposed [22]. Extension to time-delay systems have also been proposed [23]. The efficiency of the interval observer design, even with chaotic systems has been developed and applied considering parameters uncertainties of the system and biased output [108], [105].

The combination of the observers has also been improved in the case where various types of interval observers are run in parallel in a so-called "bundle of observers" [73]. These algorithms have been improved by the estimation of the observer gain providing the best estimate [40], [21]. The approach has been applied to estimation of the microalgae growth and lipid production [101].

These works are done in collaboration with Frédéric Mazenc (DISCO, Inria) and Marcelo Moisan (EMEL S.A., Chile).

6.1.4. Metabolic and genomic models

Participants: Jean-Luc Gouzé, Madalena Chaves, Alfonso Carta, Ismail Belgacem, Xiao Dong Li, Olivier Bernard, Wassim Abou-Jaoudé, Luis Casaccia, Caroline Baroukh, Rafael Muñoz-Tamayo, Jean-Philippe Steyer.

Multistability and oscillations in genetic control of metabolism

Genetic feedback is one of the mechanisms that enables metabolic adaptations to environmental changes. The stable equilibria of these feedback circuits determine the observable metabolic phenotypes. Together with D. Oyarzun from Imperial College, we considered an unbranched metabolic network with one metabolite acting as a global regulator of enzyme expression. Under switch-like regulation and exploiting the time scale separation between metabolic and genetic dynamics, we developed geometric criteria to characterize the equilibria of a given network. These results can be used to detect mono- and bistability in terms of the gene regulation parameters for any combination of activation and repression loops. We also find that metabolic oscillations can emerge in the case of operon-controlled networks; further analysis reveals how nutrient-induced bistability and oscillations can emerge as a consequence of the transcriptional feedback [27].

Global stability for metabolic models and unreduced Michaelis-Menten equations

We are interested in the uniqueness and stability of the equilibrium of reversible metabolic models. For biologists, it seems clear that realistic metabolic systems have a single stable equilibrium. However, it is known that some types of metabolic models can have no or multiple equilibria. We have made some contribution to this problem, in the case of a totally reversible enzymatic system. We prove that the equilibrium is globally asymptotically stable if it exists; we give conditions for existence and behavior in a more general genetic-metabolic loop [26]. Moreover, with the same techniques, we studied full (i.e. not reduced by any time-scale argument) Michelis-Menten reactions or chains of reactions: we prove global stability when the equilibrium exists, and show that it may not exist. This fact has important consequences for reduction of metabolic systems in a coupled genetic-metabolic system [34], [45], [70].

Interconnections of Boolean modules: asymptotic and transient behavior

A biological network can be schematically described as an input/output Boolean module: that is, both the states, the outputs, and the inputs are Boolean. The dynamics of a Boolean network can be represented by an asynchronous transition graph, whose attractors describe the system's asymptotic behavior. We have shown that the attractors of the feedback interconnection of two Boolean modules can be fully identified in terms of cross-products of the semi-attractors (states of the attractor with same output) of each module. In [82], the *asymptotic graph* was proposed, which is quite fast to compute and identifies all attractors of the interconnected system, but may also generate some spurious attractors. In [31] the *cross graph* is proposed, which exactly identifies the attractors of the interconnected system but is slower to compute. The asymptotic dynamics of high-dimensional biological networks can thus be predicted through the computation of the dynamics of two isolated smaller subnetworks. An application is, for instance, to interconnect four individual "cells" to obtain all the attractors of the segment polarity genes model in *Drosophila*.

Probabilistic approach for predicting periodic orbits in piecewise affine differential models

The state space of a piecewise affine system is partitioned into hyperrectangles which can be represented as nodes in a directed graph, so that the system's trajectories follow a path in a transition graph. Using this property we defined a *transition probability* between two nodes A and B of the graph, based on the volume of the initial conditions on the hyperrectangle A whose trajectories cross to B [15]. The parameters of the system can thus be compared to the observed or experimental transitions between two hyperrectangles. This definition is useful to identify sets of parameters for which the system yields a desired periodic orbit with a high probability, or to predict the most likely periodic orbit given a set of parameters, as illustrated by a gene regulatory system composed of two intertwined negative loops.

Structure estimation for unate Boolean models of gene regulation networks

Estimation or identification of the network of interactions among a group of genes is a recurrent problem in the biological sciences. Together with collaborators from the University of Stuttgart, we have worked on the reconstruction of the interaction structure of a gene regulation network from qualitative data in a Boolean framework. The idea is to restrict the search space to the class of unate functions. Using sign-representations, the problem of exploring this reduced search space is transformed into a convex feasibility problem. The sign-representation furthermore allows to incorporate robustness considerations and gives rise to a new measure which can be used to further reduce the uncertainties. The proposed methodology is demonstrated with a Boolean apoptosis signaling model [35].

E. coli modeling and control

In the framework of ANR project Gemco, we developed and analyzed a model of a minimal synthetic gene circuit, that describes part of the gene expression machinery in *Escherichia coli*, and enables the control of the growth rate of the cells during the exponential phase.

This model is a piecewise non-linear system with two variables (the concentrations of two gene products) and an input (an inducer). We studied the qualitative dynamics of the model and the bifurcation diagram with respect to the input. Moreover, an analytic expression of the growth rate during the exponential phase as function of the input has been derived. A relevant problem was that of parameters identifiability of this expression supposing noisy measurements of exponential growth rate. We presented such an identifiability study that we validated in silico with synthetic measurements [36].

We also studied a model of the global cellular machinery designed by D. Ropers and collaborators (IBIS team, Grenoble). This model has 11 variables and many parameters ; we explored different techniques for reduction and simplification [56], [57].

Transition graph and dynamical behavior of piecewise affine systems

We investigated the links between the topology of the transition graph and the number and stability of limit cycles in a class of two-dimensional piecewise affine biological models. To derive these structure-to-dynamics principles, we use the properties of continuity, monotonicity and concavity of Poincare maps associated with transition cycles of the transition graph [64].

Robust estimation for a hybrid model of genetic networks

State estimation problems with Boolean measurements for a classical negative loop genetic network governed by a piecewise affine (PWA) model have been studied in [39]. Observers are proposed for the cases where either full state or only partial state Boolean measurements are available. In the first case, sliding modes may occur, which leads to finite time convergence for the observer. In the second case, an algebraic computation is proposed to solve the initial condition inverse problem. The robustness of the observer for a parametric uncertain model is investigated, and we show that the error bound is proportional to the magnitude of the uncertainty.

Modeling the metabolic network in non balanced growth conditions

We have developed a new approach to represent the metabolic network of organisms for which the hypothesis of balanced growth is not satisfied [67]. This is especially true for microalgae which store carbon during the day and nitrogen during the night [44]. The proposed formalism is based on the assumption that some parts of the metabolic network satisfy the balance growth conditions, *i.e.* there is no accumulation of intermediate compounds. This hypothesis specifically applies to the main functions in the cell (respiration, photophosphorylation,...). Between two functions, some compounds can accumulate with storage/reuse kinetics. The resulting system is thus a slow-fast system.

6.2. Fields of application

6.2.1. Bioenergy

6.2.1.1. Modeling of microalgae production

Participants: Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Philipp Hartmann, Rafael Muñoz-Tamayo, Ghjuvan Grimaud, Charlotte Combe, Hubert Bonnefond, Jean-Philippe Steyer, Francis Mairet.

Experimental developments

Experiments have been carried out to study the effects of nitrogen limitation on the lipid production in microalgae [18], [17], [48] and support model development. These experiments have been carried out in the Lagrangian simulator (SEMPO), under constant or periodic light and temperature, varying the total amount of light dose in the day. The response in terms of storage carbon (triglycerides and carbohydrates) has been observed.

Other experiments were carried out to reproduce the light percept by a cell in a raceway [58]. An electronic platform was developed to reproduce the flashing light which, from the hydrodynamical studies, is likely to happen in a raceway at the cell scale. The experiments show that the microalgae adapt their pigments to the average light that they have received.

The effect in the cell cycle of both the light periodic signal and a nitrogen limitation were studied. The strong interactions of the interactions between the different phases of the cell cycle through checkpoints was highlighted [24].

This work is done in collaboration with Amélie Talec, Thomas Lacour, and Christophe Mocquet (CNRS-Océanographic Laboratory of Villefranche-sur-Mer).

Modeling the effect of temperature

The effect of temperature on microalgae has been represented by adapting the CTMI model developed for bacteria [115]. The proposed model [14], [28], was able to correctly represent the growth response to temperature for 15 different species. A procedure for model calibration and estimation of the parameter uncertainties was specially developed, allowing to gather experimental data from various sources. It was shown that different strains of the same species have a very similar response to temperature fluctuations. Moreover, for low light intensities, a simple model can represent both effects of light and temperature [14].

Modeling light distribution within a photobioreactor

The light distribution within a photobioreactor was estimated thanks to a multi photon Monte-Carlo simulation. From measurements of absorption and scattering properties, it was thus possible to extrapolate and validate the light distribution within a photobioreactor or a raceway.

Modeling lipid accumulation

We have proposed a new model for lipid production by microalgae which describes the fate of the CO_2 incorporated during photosynthesis [10]. This model describes the accumulation of neutral lipids (which can be turned into biofuel), carbohydrates and structural carbon. It has been calibrated and validated with experimental data. This model highlights and explains the phenomenon of hysteresis in lipid production which has been experimentally verified. It has been extended to account for light/dark cycles [96].

Modeling a microalgae production process

The integration of different models developed in the group [72], [96], [10],[14] was performed to represent the dynamics of microalgae growth and lipid production in raceway systems, on the basis of the dynamical model developed to describe microalgal growth in a photobioreactor under light and nitrogen limitations [72]. The strength of this model is that it takes into account the strong interactions between the biological phenomena (effects of light and nitrogen on growth, photoacclimation ...), temperature effect and the radiative transfer in the culture (light attenuation due to the microalgae).
Using these approaches, we have developed a model which predicts lipid production in raceway systems under varying light, nutrients and temperature [109], [110]. This model is used to predict lipid production in the perspective of large scale biofuel production.

6.2.1.2. Coupling growth of microalgae with hydrodynamics

Participants: Olivier Bernard, Antoine Sciandra, Philipp Hartmann, Charlotte Combe.

Modeling the coupling between hydrodynamics and biology

In collaboration with the Inria ANGE team, a model coupling the hydrodynamics of the raceway (based on multilayer Saint Venant system) with microalgae growth was developed [13]. This model is supported by the work of ANGE aiming at reproducing the hydrodynamics of the raceway, with a specific attention to the effect of the paddle wheel on the fluid.

Modeling the photosynthesis response to fast fluctuating light

The impact of the hydrodynamics on the light percept by a single cell was studied thanks to fluid dynamics simulations of a raceway pond [37] [92]. The light signals that a cell experiences at the Lagrangian scale, depending on the fluid velocity, were then estimated. A Droop-Han model was used to assess the impact of light fluctuation on photosynthesis. A new model accounting for photoacclimation was also proposed [93].

6.2.1.3. Optimization of microalgae production

Participants: Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Rafael Muñoz-Tamayo.

Numerical optimization

Using the detailed model for raceway systems, we assessed strategies for optimal operation in continuous mode [109]. Two strategies were developed. The first one resides in solving numerically an optimal control problem in which the input flow rate of the raceway is calculated such that the productivity in microalgae biomass is maximized on a finite time horizon. In the second strategy, we aimed at translating the optimization problem into a regulation problem. We proposed a simple operational criterion that when integrated in a strategy of closed-loop control allows to reach biomass productivities very near to the productivities obtained with the optimal control. We demonstrated that the practical advantages for real implementation makes our proposed controller a suitable control strategy for optimizing microalgae production in raceways.

Analytical optimization

Optimization strategies were based on simple microalgae models : first, biomass production has been optimized in a constant light environment [104], yielding results emphasizing the importance of the optical depth of the reactor. In a second work, we focused on the optimal operating conditions for the biomass productivity under day/night cycles using Pontryiagin's maximum principle (assuming a periodic working mode) [61] [90].

6.2.2. CO_2 fixation by microalgae

Participants: Olivier Bernard, Antoine Sciandra, Ghjuvan Grimaud.

Experimental work

We have run experiments to observe the response of a population of microalgal cells to various periodic light/dark or nitrate signals. The measurements show the synchronicity of the cells for some conditions. These experiments support the hypothesis that uptake of nitrogen stops during cell division [24].

Modeling cell cycle

On this basis, we have developed a structured model representing the development of microalgal cells through three main phases of their cell cycle: G1, G2 and M. The model is made of three interdependent Droop models [107]. The model was validated through extensive comparison with experimental results in both condition of periodic light forcing and nitrogen limitation. The model turns out to accurately reproduce the experimental observations [107].

Calcification of coccolithophorids

The effect of CO_2 partial pressure increase on photosynthesis and calcification of the calcareous microalgae *Emiliania huxleyi* have been experimentally observed. It results in an increase of the coccolith size together with a decrease in the calcification rate [25].

Three models accounting for the possible coupling between photosynthesis and calcification [75] were included in an ocean model; they account for settling and predation by grazers, and a bloom of coccolithophorids was simulated [76], [77].

Nitrogen fixation by nitrogenotrophs

The fixation of nitrogen by *Croccosphera watsonii* was represented with a macro metabolic model [60] [89]. The main fluxes of carbon and nitrogen are represented in the cell. The accumulation of starch during the day to fuel the nitrogenase working in the absence of oxygen during the night was the key process to explain the nitrogen fixation. The strong influence of the cell cycle was also included in the model. Finally, the model was calibrated and validated with the data of 3 experiments carried out with different duration of the light period and daily dose. The model succeded to efficiently reproduce the experimental data.

This work is done in collaboration with Sophie Rabouille (CNRS-Océanographic Laboratory of Villefranche-sur-Mer).

Including phytoplankton photoadaptation into biogeochemical models

The complexity of the marine ecosystem models and the representation of biological processes, such as photoadaptation, remain an open question. We compared several marine ecosystem models with increasing complexity in the phytoplankton physiology representation in order to assess the consequences of the complexity of photoadaptation models in biogeochemical model predictions. Three models of increasing complexity were considered, and the models were calibrated to reproduce ocean data acquired at the Bermuda Atlantic Time-series Study (BATS) from in situ JGOFS (Joint Global Ocean Flux Study) data. It turns out that the more complex models are trickier to calibrate and that intermediate complexity models, with an adapted calibration procedure, have a better prediction capability [12], [43], [42].

This work is done in collaboration with Sakina Ayata (UPMC-Océanographic Laboratory of Villefranche-sur-Mer).

6.2.3. Design of ecologically friendly plant production systems

6.2.3.1. Controlling plant pests

Participants: Frédéric Grognard, Ludovic Mailleret, Mickaël Teixeira-Alves, Nicolas Bajeux.

Optimization of biological control agent introductions

The question of how many and how frequently natural enemies should be introduced into crops to most efficiently fight a pest species is an important issue of integrated pest management. The topic of natural enemies introductions optimization has been investigated for several years [9] [111]. It had allowed to unveil the crucial influence of within-predator density dependent processes, and especially negative density dependence. In particular, we concluded that pest control is more efficiently achieved through the frequent introduction of small populations of natural enemies as compared to larger and rarer ones. Because contrarily to predatory biocontrol agents, parasitoids may be more prone to exhibit positive density dependent dynamics rather than negative ones, the current modeling effort concentrates on studying the impact of positive predator-predator interactions on the optimal introduction strategies [55].

Connected experimental research is also being pursued in the laboratory on *trichogramma spp.* which tends to show positive density dependence because of demographic stochasticity [32], and the PhD thesis of Thibaut Morel Journel (UMR ISA) has just started on this topic.

Food source diversity and classical biological control efficiency using generalist natural enemies

Because generalist biocontrol agents can feed on different food sources like, e.g. a given pest and pollen, they are capable of surviving pest absence within crops [118]. From the biological control point of view, this makes it possible to sustain natural enemies populations able to fight pests at the onset of pest attacks. Moreover, when supplied with different food types, generalists organisms are expected to thrive. Alternative prey, banker plants or more generally habitat enhancement based biological control strategies are thus becoming popular IPM (Integrated Pest Management) methods [112]. Although it has clear advantages, the simultaneous presence of various food sources also has important drawbacks: feeding on different food sources means that a given individual cannot feed on each food source at the same moment. This distraction effect thus potentially reduces the overall predation pressure imposed by the natural enemy population, and the interaction between the demographic response of the predator population and individual behavior is complex. To investigate such questions, we developed and analyzed behavioral-demographic population models taking into account the negative density dependent character of most generalist biocontrol agents. We found out that predator distraction effects can dominate the demographic response of the predator populations, potentially disrupting pest control [120]. An additional conclusion of our study, is that higher predator densities can actually bring about lesser pest suppression. Such results question current biological control practices, and show that, counter-intuitively, recording a lot of predators within fields does not ensure efficient pest control.

Plant compensation, pest control and plant-pest dynamics

Plant compensation is the process by which plants respond positively to recover from the effects of pest injury on plant growth. It is a common phenomenon, which has been repeatedly reported in various plant taxa during the last thirty years. Of special interest is the overcompensation phenomenon: consecutively to a pest attack, a plant may reach a higher biomass or have a better fitness compared to the no-pest-attack situation [65]. Although this phenomenon has mainly been documented in wild plants [65] it has also been observed on agricultural plants [121], [113]. To understand better this plant-herbivore interaction and to assess the efficacy of different pest control strategies we built a plant-pest model of plant compensatory growth. We have shown that depending on plants and pests characteristics, plant overcompensation may or may not happen. Moreover, because the model undergoes a backward bifurcation, it is shown that plant overcompensation is also dependent on the level of pest attacks and does not necessarily show up even when the plant-pest couple do have the potential to produce overcompensation [38].

This work is part of the PhD thesis of Audrey Lebon (Cirad), and done in collaboration with Yves Dumont (Cirad).

6.2.3.2. Controlling plant pathogens

Participants: Frédéric Grognard, Ludovic Mailleret, Elsa Rousseau.

Sustainable management of plant resistance

The introduction of plant strains that are resistant to one pathogen often leads to the appearance of virulent pathogenic strains that are capable of infecting the resistant plants. The resistance strain then becomes useless. It is therefore necessary to develop ways of introducing such resistance into crop production without jeopardizing its future efficiency. We did so by choosing the proportion of resistant plants that are mixed with the non-resistant ones. We studied a vector borne pathogen in a seasonal environment, with healthy crop being planted at the beginning of each season and cropped at its end, the pathogen surviving in the environment during the 'winter'. Two strategies have been proposed, one that aims at minimizing the cumulated damage over a 15 years horizon and one that aims at preventing the virulent strain outbreak. We showed that pathogen's fitness cost associated with resistance breakdown was one of the main factors governing damage reduction at the landscape scale, although the optimal resistance deployment also strongly relied on epidemic characteristics and landscape coonnectivity [16], [51]. The capacity for a virulent virus to establish itself in such an environment, as well as the evolution of the virus characteristics have been studied [59].

This work is done in collaboration with Frédéric Fabre and Benoit Moury (INRA Avignon).

Eco-evolutionary dynamics of plant pathogens in seasonal environments

The coexistence of closely related plant parasites is widespread. Yet, understanding the ecological determinants of evolutionary divergence in plant parasites remains an issue. Niche differentiation through resource specialization has been widely researched, but it hardly explains the coexistence of parasites exploiting the same host plant. Most agricultural systems in temperate environments are characterized by the cyclical presence and absence of the crop, due to cropping practices such as harvest and planting. The seasonal character of agrosystems can induce complex plant-pathogens dynamics [19] and is an important force promoting evolutionary diversification of plant pathogens [91]. Plant parasites reproduction mode may strongly interact with seasonality. In this context, we investigated the influence of cyclical parthenogenesis, i.e. the alternation of sexual and asexual reproduction phases, on the eco-evolutionary dynamics of plant parasites [80]. By means of a theoretical approach, we show that an obligate sexual event prior to overseasoning promotes evolutionary divergence in terms of investment into asexual reproduction in plant parasites. Yet, polymorphism may be transient; namely, morphs mostly investing into sexual reproduction may eventually exclude morphs mostly investing into asexual reproduction. Our findings nicely echo with recent population genetics results on *Leptosphaeria maculans*, the causal agent of the blackleg disease of canola, reporting differential investments into sexual and asexual reproduction both at the global and continental scales.

This work is part of the PhD thesis of Magda Castel (Agrocampus Ouest) and is done in collaboration with Frédéric Hamelin (Agrocampus Ouest).

6.2.4. Biological depollution - Anaerobic digestion

6.2.4.1. Coupling microalgae to anaerobic digestion

Participants: Olivier Bernard, Antoine Sciandra, Jean-Philippe Steyer, Frédéric Grognard, Philipp Hartmann.

The coupling between a microalgal pond and an anaerobic digester is a promising alternative for sustainable energy production and wastewater treatment by transforming carbon dioxide into methane using light energy. The ANR Symbiose project is aiming at evaluating the potential of this process [117], [116].

In a first stage, we developed models for anaerobic digestion of microalgae. Two approaches were used: First, a dynamic model has been developed trying to keep a low level of complexity so that it can be mathematically tractable for optimization [97], [79], [20]. Considering three main reactions, this model fits adequately the experimental data of an anaerobic digester fed with *Chlorella vulgaris* (data from INRA LBE). On the other hand, we have tested the ability of ADM1 [119] (a reference model which considers 19 biochemical reactions) to represent the same dataset. This model, after modification of the hydrolysis step [99], [100], [98] has then been used to evaluate process performances (methane yield, productivity...) and stability though numerical simulations.

6.2.4.2. Life Cycle Assessment of microalgae production

Participants: Olivier Bernard, Jean-Philippe Steyer.

This work is the result of a collaboration with Laurent Lardon and Arnaud Helias of INRA-LBE through the co-supervision of Pierre Collet's PhD thesis [83].

An analysis of the potential environmental impacts of biodiesel production from microalgae has been carried out using the life cycle assessment (LCA) methodology [94]. This study has allowed to identify the obstacles and limitations which should receive specific research efforts to make this process environmentally sustainable. This study has been updated and the effects of technological improvements (leading to higher productivities) have been compared to the source of electricity. It turns out that the overall environmental balance can much more easily be improved when renewable electricty is produced on the plant [86], [85]. As a consequence, a new paradigm to transform solar energy (in the large) into transportation biofuel is proposed, including a simultaneous energy production stage.

A LCA has been carried out to assess the environmental impact of methane production by coupling microalgae and anaerobic digestion. The study highlights the limitation derived by the low biodegradability of the considered microalgae [84] which induces a large digester design and thus more energy to mix and heat it.

6.2.5. Models of ecosystems

6.2.5.1. Optimality in consumer-resource dynamics

Participants: Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard.

Adaptive behavior in seasonal consumer-resource dynamics

In this work we studied the evolution of a consumer-resource (or predator-prey) system with seasonal character of the dynamics. We specified two main parts of the process. First, we considered the system during one season with a fixed length: the prey lay eggs continuously and the predators lay eggs or hunt the prey (choose their behavior) according to the solution of an optimal control problem [66]. We then examined how (resident) predators adopting this optimal behavior would fare when faced with a small population of selfish mutants that would be identical to the resident but would have the freedom to choose a different behavior. We studied the resulting optimal control problem where the mutants maximize their own number of offspring using the knowledge of the resident's behavior, and showed that, in most situations, mutants can take advantage of their low frequency and fare better than the residents. Over the course of a large number of seasons, the mutants replace the residents, only to find themselves applying the original resident behavior [52].

Optimal foraging and residence times variations

Charnov's marginal value theorem (MVT) [81] is a central tenet of ecological theory. In fragmented environments, the MVT connects the quality and distribution of patches to the optimal time an individual should spend on any patch, and thus the rate of movement in the habitat. Unfortunately, it does not offer explicit predictions regarding how changing habitat quality would affect residence times. In this work, we answer that question in a very general setting, for habitats with homogeneous or heterogeneous patches and with general fitness functions. We then particularize it to the resource consumption framework and indicate how the residence times variations relate to the curvatures of the functional responses [49], [78].

This last work is done in collaboration with Vincent Calcagno and Eric Wajnberg (INRA Sophia Antipolis)

6.2.5.2. Growth models of zooplankton

Participants: Jean-Luc Gouzé, Jonathan Rault, Eric Benoît.

The model built to describe a zooplankton community is some variant of the McKendrick-Von Foerster Equation. The model includes cannibalism within zooplankton communities and predation on phytoplankton. Dynamic mass budget theory is used in order to describe individual behavior and allows mass conservation. Also we have added phytoplankton dynamics, and we use environmental data as an input for the model. The aim is to compare simulations with data provided by the Laboratoire d'Océanographie de Villefranche (Lars Stemmann). We have also built a discrete size-structured model. Discrete models are less numerically demanding and so can be more easily incorporated into bigger models. Moreover the study of discrete models are often easier than that of continuous ones. We focus our study on the impact of cannibalism within the zooplankton community and show that under some hypotheses, cannibalism can stabilize the equilibrium of the model [29], [11]. We also address the problem of control of such models (by harvesting or biological control); we obtain results for stabilization of the equilibrium [41], [11].

6.3. Software design

Participants: Olivier Bernard, Mélaine Gautier.

Over the years, BIOCORE has been developing a software framework for bioprocess control and supervision called ODIN [71]. This C++ application (working under Windows and Linux) enables researchers and industrials to easily develop and deploy advanced control algorithms through the use of a Scilab interpreter [46], [47]. It also contains a Scilab-based process simulator which can be harnessed for experimentation and training purposes. ODIN is primarily developed in the C++ programming language and uses CORBA to define component interfaces and provide component isolation. ODIN is a distributed platform, enabling remote monitoring of the controlled processes as well as remote data acquisition. Recently, a software development effort has been directed to the graphical user interface, a synoptic view component, new drivers for the experimental hardware and integration of the PlantML data exchange format. ODIN has been tested on four different processes and has been set up with Eric Latrille to supervise the 66m2 high rate pond at the LBE, INRA Narbonne.

BONSAI Project-Team

6. New Results

6.1. High-throughtput sequence processing

- Within the PhD of T. T. Tran, we proposed a new indexing structure adapted to GPUs. We studied an indexing scheme with perfect hashing functions, and developed a prototype written in openCL for a read mapper. This read mapper has a sensitivity comparable to state-of-the-art read mappers, and provides substantial time gains in some cases.
- Within our collaboration with the Lille hospital on the follow-up of leukemia residual disease, we proposed a new heuristic to study immunological VDJ recombinations and follow their evolution along the time. The method is under testing on several datasets obtained from the Ion Torrent sequencer at IRCL (Institut de Recherche sur le Cancer de Lille).
- Within the PhD of E. Kopylova, we designed an new algorithm to filter out ribosomal RNA sequences from RNA raw data produced in metatranscriptomic sequencing. The method combines text indexing techniques, with the Burst trie, and Universal Levenshtein automaton to allow for seraching with errors. An article has been published the journal *Bioinformatics* [4].

6.2. Noncoding RNAs

- We designed a new algorithm to produce all locally optimal secondary structures of an RNA sequence. Locally optimal secondary structures are thermodynamically stable RNA structures that are maximal for inclusion: they cannot be extended without producing a conflict between base pairs in the secondary structure, or increasing the free energy. This was published in *Journal for Computational Biology* [7].
- We took part to a collaborative work on benchmarking for RNA structure comparison. This work has been published in *Advances in bioinformatics* [2].

6.3. Genomic rearrangements

- Within the context of the PhD of A. Thomas:
  - We designed an algorithm for finding the minimal number of block interchanges required to transform a duplicated linear genome into a tandem duplicated linear genome. We provide a formula for the distance as well as a polynomial time algorithm for the sorting problem. This work was published in the conference *Bioinformatics* [13].
  - We explored a new problem concerning tandem halving, that is reconstructing a non-duplicated ancestor to a partially duplicated genome in a model where duplicated content is caused by several tandem duplications. We provide a distance in $O(n)$ time and a scenario in $O(n^2)$ time. We considered several problems related to multiple tandem reconstruction and proved that the simpliest of reconstructing 2 tandems is NP-hard. This work was published in the conference WABI 2012 [14].
- In the context of ancestral genome reconstruction, we designed an algorithm for the identification of Minimal Conflicting Sets (MCS) rows in a biological binary matrix. We provided a $O(n^2m^2 + nm^7)$ time algorithm, largely improving the up-to-date best algorithm in $O(m^6n^5(m+n)^2log(m+n))$ time. This work was published in the conference CPM 2012 [11].

- In the context of the comparison of sets of alternative gene transcripts, we designed a general framework to compare sets of transcripts that are transcribed from orthologous loci of several species. The model is based on the construction of a common reference sequence, and on annotations that allow the reconstruction of ancestral sequences, the identification of conserved events, and the inference of gains and losses of donor/acceptors sites, exons, introns and transcripts. This work was published in the conference ISBRA 2012 [12].

6.4. Nonribosomal peptides

- With the arrival of Ammar Hasan, a postdoc researcher, we started a new project on the prediction of nonribosomal peptides activity. We defined a novel peptide fingerprint based on monomer composition. This fingerprints is used for peptide similarity searching and for activity prediction. This work was published in *Journal of Computer-Aided Molecular Design* [1].
- We participated in the writing of a review dedicated to kurstakin, a nonribosomal lipopeptide synthetized by several Bacillus genus and published in *Applied microbiology and biotechnology* [3].
- The collaboration with members of EPI Orpailleur (CRI Nancy) succeeded in designing a protocol to discover new nonribosomal peptide synthetases in bacterial genomes and then annotate them in order to predict the peptide they produce. It was published in JOBIM 2012 [16].

CARMEN Team

5. New Results

5.1. Models

- [12]: we explain the links between the solutions of the bidomain and monodomain models using some analytical arguments. The result is partially based on the theory of the bidomain operator explained in [11].
- [23]: Fibre structure and anisotropy is a determinant issue to provide accurate simulations of the electrical activity of atrial tissue. Though, atrial fibre architecture remains unreachable to standard imagery techniques on patients. A method to construct models of the fibre architecture on patient-specific geometries is then a key for numerical simulations of atrial tissues. Such a method is proposed. Pathological and non pathological patient specific surface models of the left atria (LA) are defined. Hence, a pathological scenario is explored : a mechanism of micro- reentry in the left superior pulmonary vein (LSPV) and its interaction with the sinus rhythm (SR).

5.2. Numerical techniques

- [19]: In this paper we propose a preconditioning for the bidomain model either for an isolated heart or in an extended framework including a coupling with the surrounding tissues (the torso). The preconditioning is based on a formulation of the discrete problem that is shown to be symmetric positive semi-definite. A block LU decomposition of the system together with a heuristic approximation (referred to as the monodomain approximation) are the key ingredients for the preconditioning definition. Numerical results are provided for two test cases: a 2D test case on a realistic slice of the thorax based on a segmented heart medical image geometry, a 3D test case involving a small cubic slab of tissue with orthotropic anisotropy. The analysis of the resulting computational cost (both in terms of CPU time and of iteration number) shows an almost linear complexity with the problem size, i.e. of type nlogα(n) (for some constant α) which is optimal complexity for such problems.

5.3. Medical applications of numerical models

- [26]: We computed some bidomain solutions for use by M. Pop and M. Sermesant in the STACOM'11 challenge from the MICCAI 2011 conference and derived collaborative article [26].
- [18]: The aim of this study was to describe a new familial cardiac phenotype and to elucidate the electrophysiological mechanism responsible for the disease. Mutations in several genes encoding ion channels, especially SCN5A, have emerged as the basis for a variety of inherited cardiac arrhythmias. Three unrelated families comprising 21 individuals affected by multifocal ectopic purkinje-related premature contractions (MEPPC) characterized by narrow junctional and rare sinus beats competing with numerous premature ventricular contractions with right and/or left bundle branch block patterns were identified. All the affected subjects carried the same transition in the SCN5A gene. Patch-clamp studies revealed a net gain of function of the sodium channel, leading, in silico, to incomplete repolarization in Purkinje cells responsible for premature ventricular action potentials. In vitro and in silico studies recapitulated the normalization of the ventricular action potentials in the presence of quinidine.
- [22]: In some cases, the standard methods to construct activation maps based on the derivatives of the signals may lead to inaccurate results. In this paper, we evaluated a novel Directional Activation Algorithm (DAA) based on EGM analysis. The DAA calculates the time delays between adjacent EGMs and assigns to each a localized propagation vector. The accuracy of the proposed methodology is compared with known activities obtained from a monodomain, isotrope, Beeler-Reuter model of the atria.

- [20]: Although the ECG is a widely used tool, the ionic basis underlying its changes caused by drugs and diseases are often unclear. In this work we present a computational model of the human ECG capable of representing drug-induced effects from the ionic to the surface potential level. We use the state-of-the-art bidomain model coupled to a membrane kinetics model in the heart and the Laplace equation in the torso. The membrane kinetics are represented by a detailed physiological human action potential model. We modified the potassium (respectively sodium) representation in the model in order to introduce the ion channel/drug interactions representing classIII (respectively class I) drugs. The drug model is represented by an ion channel conduction block depending on the IC50 value and the drug dose. We conduct numerical simulation of the ECGs measured on the surface of the thorax and could assess each of the potassium and sodium block effects (for class I and class III drugs).

5.4. Inverse problems

- [24]: The treatment of atrial fibrillation has greatly changed in the past decade. Ablation therapy, in particular pul- monary vein ablation, has quickly evolved. However, the sites of the trigger remain very difficult to localize. In this study we propose a machine-learning method able to non-invasively estimate a single site trigger. The machine learning technique is based on a kernel ridge regression algorithm. In this study the method is tested on a simulated data. We use the monodomain model in order to simulate the electrical activation in the atria. The ECGs are computed on the body surface by solving the Laplace equation in the torso.
- [16]: In the present paper, an optimal control problem constrained by the tridomain equations in electrocardiology is investigated. The state equations consisting in a coupled reaction–diffusion system modeling the propagation of the intracellular and extracellular electrical potentials, and ionic currents, are extended to further consider the effect of an external bathing medium. The existence and uniqueness of solution for the tridomain problem and the related control problem is assessed, and the primal and dual problems are discretized using a finite volume method which is proved to converge to the corresponding weak solution. In order to illustrate the control of the electrophysiological dynamics, we present some preliminary numerical experiments using an efficient implementation of the proposed scheme.
- [17]: This note is devoted to the analysis of the null controllability of a nonlinear reaction–diffusion system, approximating a parabolic–elliptic system, modeling electrical activity in the heart. The uniform, with respect to the degenerating parameter, null controllability of the approximating system by a single control force acting on a subdomain is shown. The proof needs a precise estimate with respect to the degenerating parameter and it is done combining Carleman estimates and energy inequalities.

CLIME Project-Team

6. New Results

6.1. New methods for data assimilation

Since the beginning, Clime is focused on developing new techniques for data assimilation in geophysical sciences. Clime is active on several of the most challenging theoretical aspects of data assimilation: data assimilation methods based on non-Gaussian assumptions, methods for estimating errors, ensemble filtering techniques, 4D variational assimilation approaches, ensemble-variational methods, etc. This year, we revisited several of these topics. A dual algorithm has been developed for the finite-size ensemble Kalman filter, that shows how to estimate optimal inflation that counteracts sampling errors. A variational method coupled to a subgrid scale statistical model has been introduced and validated to quantify the representativeness errors. We also started to work on ensemble variational methods that are brand new techniques emerging in the meteorological data assimilation field.

6.1.1. Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems

Participants: Marc Bocquet, Pavel Sakov [NERSC, Norway].

The finite-size ensemble Kalman filter (EnKF-N) is an ensemble Kalman filter (EnKF) which, in perfect model condition, does not require inflation because it partially accounts for the ensemble sampling errors. For the Lorenz '63 and '95 toy-models, it was so far shown to perform as well or better than the EnKF with an optimally tuned inflation. The iterative ensemble Kalman filter (IEnKF) is an EnKF which was shown to perform much better than the EnKF in strongly nonlinear conditions, such as with the Lorenz '63 and '95 models, at the cost of iteratively updating the trajectories of the ensemble members. This study aims at further exploring the two filters, and at combining both into an EnKF that does not require inflation in perfect model condition and which is as efficient as the IEnKF in very nonlinear conditions.

In this study EnKF-N is first introduced and a new implementation is developed. It decomposes EnKF-N into a cheap two-step algorithm that amounts to computing an optimal inflation factor. This offers a justification of the use of the inflation technique in the traditional EnKF and why it can often be efficient. Secondly, the IEnKF is introduced following a new implementation based on the Levenberg-Marquardt optimization algorithm. Then, the two approaches are combined to obtain the finite-size iterative ensemble Kalman filter (IEnKF-N). Several numerical experiments are performed on IEnKF-N with the Lorenz '95 model. These experiments demonstrate its numerical efficiency as well as its performance that offer, at least, the best of both filters.

6.1.2. Accounting for representativeness errors in the inversion of atmospheric constituent emissions: Application to the retrieval of regional carbon monoxide fluxes

Participants: Mohammad Reza Koohkan, Marc Bocquet.

A four-dimensional variational data assimilation system (4D-Var) is developed to retrieve carbon monoxide (CO) fluxes at regional scale, using an air quality network. The air quality stations that monitor CO are proximity stations located close to industrial, urban or traffic sources. The mismatch between the coarsely discretized Eulerian transport model and the observations, inferred to be mainly due to representativeness errors in this context, leads to a bias (averaged simulated concentrations minus observed concentrations) of the same order of magnitude as the concentrations. 4D-Var leads to a mild improvement in the bias because it does not adequately handle the representativeness issue. For this reason, a simple statistical subgrid model is introduced and is coupled to 4D-Var. In addition to CO fluxes, the optimization seeks to jointly retrieve *influence coefficients*, which quantify each station's representativeness. The method leads to a much better representation of the CO concentration variability, with a significant improvement of statistical indicators.

The resulting increase in the total inventory estimate is close to the one obtained from remote sensing data assimilation. This methodology and experiments suggest that information useful at coarse scales can be better extracted from atmospheric constituent observations strongly impacted by representativeness errors.

6.1.3. Real-time data assimilation

Participants: Vivien Mallet, Anne Tilloy, Fabien Brocheton [Numtech], David Poulet [Numtech], Cécile Honoré [Airparif], Édouard Debry [INERIS].

Based on Verdandi, Polyphemus and the "Urban Air Quality Analysis" software, real-time data assimilation was carried out at urban scale. The Best Linear Unbiased Estimator (BLUE) was computed for every hourly concentration map that the ADMS model computed. A posteriori tests were conducted over Clermont-Ferrand and Paris. We addressed the key issue of the covariance of the state error. The form of the error covariance between two points was determined based on the road network, considering the distance between points along the road and the distance of each point to the road. A few parameters (primarily two decorrelation lengths) were determined thanks to cross validation with several months of simulations and observations. The results showed strong improvements even at locations where no data was assimilated.

At larger scale, the data assimilation library Verdandi was used to apply data assimilation (optimal interpolation) with the air quality model Chimere. This preliminary work will help INERIS to apply optimal interpolation for ozone and particulate matter in the operational platform Prev'air.

6.2. Inverse modeling

Many of this year's studies have focused on inverse modeling, including the reconstruction of the Fukushima radionuclide atmospheric and marine source terms. All were targeted to a particular application. However most of them include new methodological developments, in particular non-Gaussian data assimilation schemes.

6.2.1. Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: Application to the reconstruction of the Fukushima Daiichi source term

Participants: Victor Winiarek, Marc Bocquet, Olivier Saunier [IRSN], Anne Mathieu [IRSN].

The aim of this research activity is the implementation of data assimilation methods, particularly inverse modeling methods, in the context of an accidental radiological release from a nuclear power plant and their application in the specific case of the Fukushima Daiichi accident. The particular methodological focus is the a posteriori estimation of the prior errors statistics. In the case of the Fukushima Daiichi accident, the number of available observations is small compared to the number of source parameters to retrieve and the reconstructed source is highly sensitive to the prior errors. That is the why they need to be well established and justified. In this aim, three methods have been proposed: one method relies on a L-curve estimation technique, another one on the Desroziers' iterative scheme and the last method, assumed to be the most robust, relies on the maximum likelihood principle, generalised to a non-Gaussian context. These three methods have been applied to the reconstruction of cesium-137 and iodine-131 source terms from the Fukushima Daiichi accident. Because of the poor observability of the Fukushima Daiichi emissions, these methods provide lower-bounds for cesium-137 and iodine-131 reconstructed activities. Nevertheless, with the new method based on semi-Gaussian statistics for the background errors, the lower-bound estimates for cesium-137, $1.2 - 4.0\,10^{16}$ Bq with an estimated standard deviation range of $15 - 20\%$, and for iodine-131, $1.9 - 3.8\,10^{17}$ Bq with an estimated standard deviation range of $5 - 10\%$, are of the same order of magnitude as those provided by the Japanese Nuclear and Industrial Safety Agency, and about 5 to 10 times less than the Chernobyl atmospheric releases.

6.2.2. Assessment of the amount of Cesium-137 released into the Pacific Ocean after the Fukushima accident and analysis of its dispersion in Japanese coastal waters

Participants: Claude Estournel [LA], Emmanuel Bosc [IAEA], Marc Bocquet, Caroline Ulses [LA], Patrick Marsailex [LA], Victor Winiarek, Iolanda Osvath [IAEA], Cyril Nguyen [LA,LEGOS], Thomas Duhaut [LA], Florent Lyard [LEGOS], Héloïse Michaud [LA], Francis Auclair [LA].

Numerical modeling was used to provide a new estimate of the amount of cesium-137 released directly into the ocean from the Fukushima Daiichi nuclear power plant (NPP) after the accident in March 2011 and to gain insights into the physical processes that led to its dispersion in the marine environment during the months following the accident. An inverse method was used to determine the time-dependent cesium-137 input responsible for the observed concentrations. The method was then validated through comparisons of the simulated concentrations with concentrations measured in seawater at different points in the neighborhood of the nuclear power plant. An underestimation was noticed for stations located 30 km offshore. The resulting bias in the release inventory was estimated. Finally, the maximum cesium-137 activity released directly to the ocean was estimated to lie between 5.1 and 5.5 PBq (Peta Becquerel = 1015 Bq) but uncertainties remain on the amount of radionuclides released during the first few days after the accident. This estimate was compared to previous ones and differences were further analysed. The temporal and spatial variations of the cesium-137 concentration present in the coastal waters were shown to be strongly related to the wind intensity and direction. During the first month after the accident, winds blowing toward the south confined the radionuclides directly released into the ocean to a narrow coastal band. Afterwards, frequent northward wind events increased the dispersion over the whole continental shelf, leading to strongly reduced concentrations.

6.2.3. What eddy-covariance measurements tell us about prior land flux errors in CO2-flux inversion schemes?

Participants: Frédéric Chevallier [LSCE], Tao Wang [LSCE], Philippe Ciais [LSCE], Marc Bocquet, Altaf Arain [McMaster University, Canada], Alessandro Cescatti [Joint Research Centre, Italy], Jiquan Chen [University of Toledo, USA], Johannes Dolman [Vrije Universiteit, the Netherlands], Beverly Law [Oregon State University, USA], Hank Margolis [Université Laval, Canada], Leonardo Montagnani [University of Bolzano, Italy].

To guide the future development of CO2-atmospheric inversion modeling systems, we analysed the errors arising from prior information about terrestrial ecosystem fluxes. We compared the surface fluxes calculated by a process-based terrestrial ecosystem model with daily averages of CO2 flux measurements at 156 sites across the world in the FLUXNET network. At the daily scale, the standard deviation of the model-data fit was 2.5 gC·m-2·d-1; temporal autocorrelations were significant at the weekly scale (> 0.3 for lags less than four weeks), while spatial correlations were confined to within the first few hundred kilometers (< 0.2 after 200 km). Separating out the plant functional types did not increase the spatial correlations, except for the deciduous broad-leaved forests. Using the statistics of the flux measurements as a proxy for the statistics of the prior flux errors was shown not to be a viable approach. A statistical model allowed us to upscale the site-level flux error statistics to the coarser spatial and temporal resolutions used in regional or global models. This approach allowed us to quantify how aggregation reduces error variances, while increasing correlations. As an example, for a typical inversion of grid point (300 km × 300 km) monthly fluxes, we found that the prior flux error follows an approximate e-folding correlation length of 500 km only, with correlations from one month to the next as large as 0.6.

6.3. Monitoring network design

In this section, we report studies that are related to the evaluation of monitoring networks and to new monitoring strategies. This year, network designs techniques have been applied to the inverse modeling of CO2 fluxes.

6.3.1. Network design for mesoscale inversions of CO2 sources and sinks

Participants: Thomas Lauvaux [Pennsylvania State University, USA], Andy Schuh [Colorado State University, USA], Marc Bocquet, Lin Wu, Scott Richardson [Pennsylvania State University, USA], Natasha Miles [Pennsylvania State University, USA], Ken Davis [Pennsylvania State University, USA].

Recent instrumental deployments of regional observation networks of atmospheric CO2 mixing ratios have been used to constrain carbon sources and sinks using inversion methodologies. In this study, we performed sensitivity experiments using observation sites from the Mid Continent Intensive experiment to evaluate the required spatial density and locations of CO2 concentration towers based on flux corrections and error reduction analysis. In addition, we investigated the impact of prior flux error structures with different correlation lengths and biome information. We show that, while the regional carbon balance converged to similar annual estimates using only two concentration towers over the region, additional sites were necessary to retrieve the spatial flux distribution of our reference case (using the entire network of eight towers). Local flux corrections required the presence of observation sites in their vicinity, suggesting that each tower was only able to retrieve major corrections within a hundred of kilometers around, despite the introduction of spatial correlation lengths (100 to 300 km) in the prior flux errors. We then quantified and evaluated the impact of the spatial correlations in the prior flux errors by estimating the improvement in the CO2 model-data mismatch of the towers not included in the inversion. The overall gain across the domain increased with the correlation length, up to 300 km, including both biome-related and non-biome-related structures. However, the spatial variability at smaller scales was not improved. We conclude that the placement of observation towers around major sources and sinks is critical for regional-scale inversions in order to obtain reliable flux distributions in space. Sparser networks seem sufficient to assess the overall regional carbon budget with the support of flux error correlations, indicating that regional signals can be recovered using hourly mixing ratios. However, the smaller spatial structures in the posterior fluxes are highly constrained by assumed prior flux error correlation lengths, with no significant improvement at only a few hundreds of kilometers away from the observation sites.

6.3.2. Potential of the International Monitoring System radionuclide network for inverse modeling

Participants: Mohammad Reza Koohkan, Marc Bocquet, Lin Wu, Monika Krysta [The Preparatory Commission for the Comprehensive Nuclear Test-Ban Treaty Organization, UNO].

The International Monitoring System (IMS) radionuclide network enforces the Comprehensive Nuclear-Test-Ban Treaty, which bans nuclear explosions. We have evaluated the potential of the IMS radionuclide network for inverse modeling of the source, whereas it is usually assessed by its detection capability. To do so, we have chosen the *degrees of freedom for the signal* (DFS), a well established criterion in remote sensing, in order to assess the performance of an inverse modeling system. Using a multiscale data assimilation technique, we have computed optimal adaptive grids of the source parameter space by maximizing the DFS. This optimization takes into account the monitoring network, the meteorology over one year (2009) and the relationships between the source parameters and the observations derived from the FLEXPART Lagrangian transport model. Areas of the domain, where the grid-cells of the optimal adaptive grid are large, emphasize zones where the retrieval is more uncertain, whereas areas, where the grid-cells are smaller and denser, stress regions where more source variables can be resolved. The observability of the globe through inverse modeling is studied in strong, realistic and small model error cases. The strong error and realistic error cases yield heterogeneous adaptive grids, indicating that information does not propagate far from the monitoring stations, whereas in the small error case, the grid is much more homogeneous.

In all cases, several specific continental regions remain poorly observed such as Africa as well as the tropics, because of the trade winds.

The northern hemisphere is better observed through inverse modeling (more than 60% of the total DFS), mostly because it contains more IMS stations. This unbalance leads to a better performance of inverse modeling in the northern hemisphere winter. The methodology is also applied to the subnetwork composed of the stations of the IMS network that measure noble gases.

6.4. Reduction and emulation

The use of environmental models raise a number of problems due to:

- the dimension of their inputs, which can easily be $10^5 - -10^8$ at every time step;

- the dimension of their state vector, which is usually $10^5 - -10^7$;
- their high computational cost.

In particular, the application of data assimilation methods and uncertainty quantification techniques may require dimension reduction and cost reduction. The dimension reduction consists in projecting the inputs and the state vector to low-dimensional subspaces. The cost reduction can be carried out by emulation, i.e., the replacement of costly components with fast surrogates.

6.4.1. Reduction and emulation of a chemistry-transport model

Participants: Vivien Mallet, Serge Guillas [University College London].

Both reduction and emulation were applied to the dynamic air quality model Polair3D from Polyphemus. The reduction relied on proper orthogonal decomposition on the input data and on the state vector. The dimension of the reduced subspace for the input data is about 80, while the dimension of the reduced state vector is less than 10. The projection of the state vector on its reduced subspace can be carried out before every integration time step, so that one can reproduce a full state trajectory (in time) using the reduced model.

Significant advances were made to emulate the reduced model, which requires about 90 inputs (reduced input data and reduced state vector) and computes about 10 outputs (reduced state vector). 90 inputs is however a large number to build an emulator using a classical approach like krigging. Promising results were however obtained with an interpolation method based on inverse distance weighting.

6.4.2. Reduction and emulation of a static air quality model

Participants: Vivien Mallet, Anne Tilloy, Fabien Brocheton [Numtech], David Poulet [Numtech].

The dimension reduction was applied to the outputs of the urban air quality model ADMS Urban, which is a static model with low-dimensional inputs and high-dimensional outputs. A proper orthogonal decomposition on the outputs allowed us to drastically reduce their dimension, from 10^4 to just a few scalars. First attempts of emulation of the reduced model rely on Gaussian process emulation.

6.4.3. Motion estimation from images with a wavelets reduced model

Participants: Giuseppe Papari, Isabelle Herlin, Etienne Huot, Karim Drifi.

The dimension reduction was applied to an image model, composed of Lagrangian constancy of velocity and transport of image brightness. Wavelets basis have been computed on the image domain for subspaces of images, motion fields and divergence-free motion fields. Image assimilation with this reduced model allows to estimate smooth velocity fields with properties defined by user. This also solves the issue of complex geographical domains and the difficulty of applying boundary conditions on these domains. First results are obtained with a reduced dimension of motion to a few scalars, to be compared with the original problem that has the size of image domain.

6.5. Ensemble forecasting with sequential aggregation

The aggregation of an ensemble of forecasts is an approach where the members of an ensemble are given a weight before every forecast time, and where the corresponding weighted linear combination of the forecasts provides an improved forecast. A robust aggregation can be carried out so as to guarantee that the aggregated forecast performs better, in the long run, than any linear combination of the ensemble members with time-independent weights. The approaches are then based on machine learning. The aggregation algorithms can be applied to forecast analyses (generated from a data assimilation system), so that the aggregated forecasts are naturally multivariate fields.

6.5.1. Application of sequential aggregation to meteorology and air quality

Participants: Anne Tilloy, Vivien Mallet, Fabien Brocheton [Numtech], David Poulet [Numtech].

Nowadays it is standard procedure to generate an ensemble of simulations for a meteorological forecast. Usually, meteorological centers produce a single forecast, out of the ensemble forecasts, computing the ensemble mean (where every model receives an equal weight). It is however possible to apply aggregation methods. Each time new observations are available, new weights for the linear combination are computed and applied for the next forecast. We applied the discounted ridge regression algorithm, which we previously introduced for sequential aggregation of air quality forecasts, to forecast wind and temperature at given observation stations. The ensemble was generated with forecasts at different range from two models. The aggregation proved to be efficient for one-day forecasts at least.

The discounted ridge regression was also applied to the simulations of the Air Quality Modeling Evaluation International Initiative (AQMEII) over Europe and North America, for different pollutants (gases and particulate matter).

6.5.2. Sequential aggregation with uncertainty estimation

Participants: Vivien Mallet, Sergiy Zhuk [IBM research], Paul Baudin, Gilles Stoltz [CLASSIC], Karine Sartelet [CEREA].

A new issue is the estimation of the uncertainties associated with the aggregated forecasts. One investigated direction relies on the framework of machine learning, with the aggregation of an ensemble of probability density functions instead of the point forecasts of the ensemble.

Another direction, which led to finalized results in 2012, is to reformulate the aggregation problem in a filtering problem for the weights. The weights are supposed to satisfy some dynamics with unknown model error, which defines the state equation of a filter. An observation equation compares the aggregated forecast with the observations (or analyses) with known observational error variance. The filter finally computes estimates for the weights and quantifies their uncertainties. We applied a Kalman filter and a minimax filter for air quality forecasting.

6.6. Image assimilation

Sequences of images, such as satellite acquisitions, display structures evolving in time. This information is recognized of major interest by forecasters (meteorologists, oceanographers, *etc*) in order to improve the information provided by numerical models. However, these satellite images are mostly assimilated in geophysical models on a point-wise basis, discarding the space-time coherence visualized by the evolution of structures such as clouds. Assimilating in an optimal way image data is of major interest and this issue should be considered in two ways:

- from the model's viewpoint, the problem is to control the location of structures using the observations,
- from the image's viewpoint, a model of the dynamics and structures has to be built from the observations.

6.6.1. Divergence-free motion estimation

Participants: Dominique Béréziat [UPMC], Isabelle Herlin, Sergiy Zhuk [IBM Research, Ireland].

This research addresses the issue of divergence-free motion estimation on an image sequence, acquired over a given temporal window. Unlike most state-of-the-art technics, which constrain the divergence to be small thanks to Tikhonov regularization terms, a method that imposes a null value of divergence of the estimated motion is defined.

Motion is characterized by its vorticity value and assumed to satisfy the Lagragian constancy hypothesis. An image model is then defined: the state vector includes the vorticity, whose evolution equation is derived from that of motion, and a pseudo-image that is transported by motion. An image assimilation method, based on the 4D-Var technics, is defined and developed that estimates motion as a compromise between the evolution equations of vorticity and pseudo-image and the observed sequence of images.

The method is applied on Sea Surface Temperature (SST) images acquired over Black Sea by NOAA-AVHRR sensors. The divergence-free assumption is roughly valid on these acquisitions, due to the small values of vertical velocity at the surface. Fig. 2 displays data and results. As no ground truth of motion is available, the method is quantified by the value of correlation between the pseudo-images and real acquisitions [28].

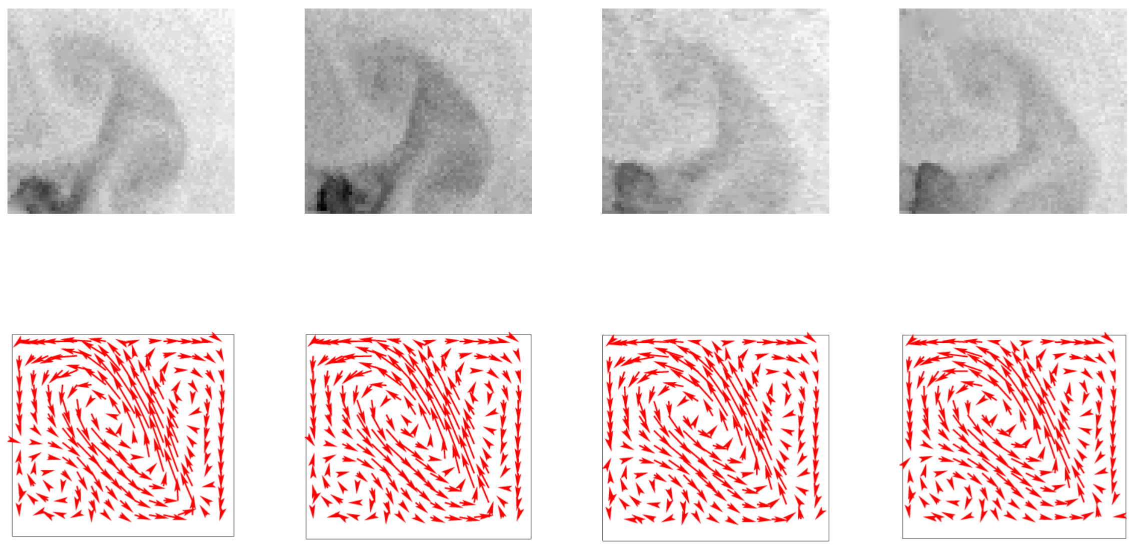


Figure 2. SST image observations and motion results.

6.6.2. Improvement of motion estimation by assessing errors on the dynamics

Participants: Dominique Béréziat [UPMC], Isabelle Herlin.

Data assimilation technics are used to retrieve motion from image sequences. These methods require a model of the underlying dynamics, displayed by the evolution of image data. In order to quantify the approximation linked to the chosen dynamic model, we consider adding a model error term in the evolution equation of motion and design a weak formulation of 4D-Var data assimilation. The cost function to be minimized simultaneously depends on the initial motion field, at the begining of the studied temporal window, and on the error value at each time step. The result allows to assess the model error and analyze its impact on motion estimation [27].

This error assessment method is evaluated and quantified on twin experiments, as no ground truth would be available for real image data. Fig. 3 shows four frames of a series of observations obtained by integrating the evolution model from an initial condition on image and velocity field (the ground truth $\mathbf{w}_{\mathrm{ref}}(0)$ displayed on the left of Fig. 4). An error value is added at each time step on the motion value, when integrating the simulation model. This error is a constant bias.

We performed two data assimilation experiments. The first one considers the evolution model as perfect, with no error in the evolution equation. It is denoted PM (for Perfect Model). The second one, denoted IM (for Imperfect Model) involves an error in the motion evolution equation. In Fig.4 are displayed the motion fields retrieved by PM and IM at the beginning of the temporal window.

As it can be seen, IM computes a correct velocity field while PM completely fails.

6.6.3. Nonlinear Observation Equation For Motion Estimation

Participants: Dominique Béréziat [UPMC], Isabelle Herlin.

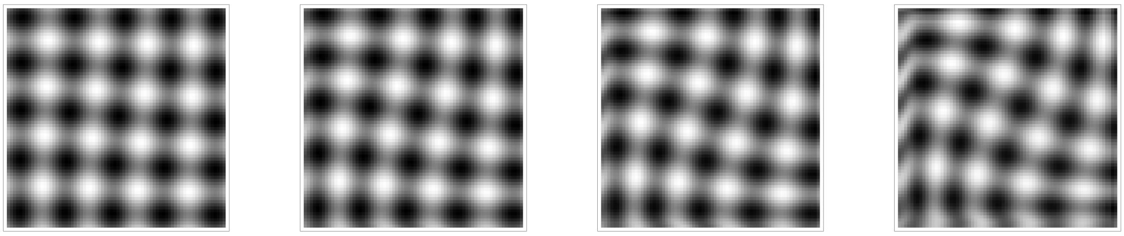


Figure 3. Observations Images.

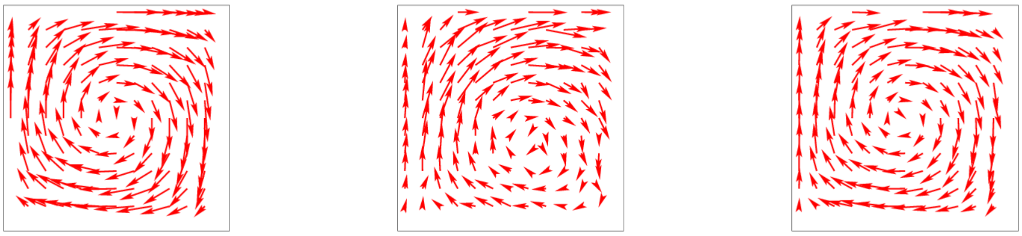


Figure 4. Left: ground-truth, middle: PM, right: IM.

In the image processing literature, the optical flow equation is usually chosen to assess motion from an image sequence. However, it corresponds to an approximation that is no more valid in case of large displacements. We evaluated improvements obtained when using the non linear transport equation of the image brightness by the velocity field [25]. A 4D-Var data assimilation method is designed that simultaneously solves the evolution equation and the observation equation, in its non linear and linearized form. The comparison of results obtained with both observation equations is quantified on synthetic data and discussed on oceanographic Sea Surface Temperature (SST) images. We show that the non linear model outperforms the linear one, which underestimates the motion norm. Fig.5 illustrates this on SST images (motion vectors are displayed by arrows).

The aim of this research is to achieve a correct estimation of motion when the object displacement is greater than its size. Howerver, in this case, coarse-to-fine incremental methods as well as the non linear data assimilation method fail to retrieve a correct value. The perspective is then to include, in the state vector, a variable describing the trajectory of pixels. The observation operator will then measure the effective displacement of pixels, according to their trajectories, and allow a better estimation of motion value.

6.6.4. Sliding windows method for motion estimation on long temporal image sequences

Participants: Karim Drifi, Isabelle Herlin.

This study concerns the estimation of motion fields from satellite images on long temporal sequences. The huge computational cost and memory required by data assimilation methods on the pixel grid makes impossible to use these techniques on long temporal intervals. For a given dynamic model (named full model), on the pixel grid, the Galerkin projection on subspaces provides a reduced model, that allows image assimilation at low cost. The definition of this reduced model however requires defining an optimal subspace of motion. A **sliding windows** method is thus designed:

- The long image sequence is split into small temporal windows that half overlap in time.

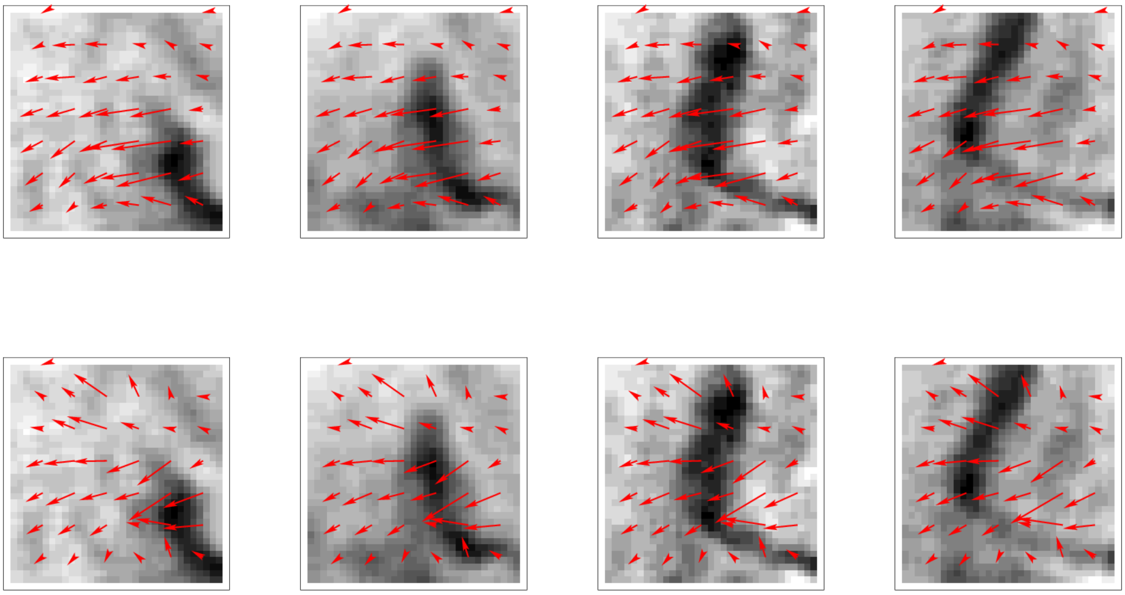


Figure 5. Top: Non-linear observation equation. Bottom: linear.

- Data assimilation in the full model is applied on the first window to retrieve the motion field.
- The estimate of motion field at the beginning of the second window makes it possible to define the subspace for motion and a reduced model is obtained by Galerkin projection.
- Data assimilation in the reduced model is applied for this second window.
- The process is then iterated for the next window until the end of the whole image sequence.

Experiments were designed to quantify the results of this sliding windows method with base obtained by Principal Orthogonal Decomposition or computed as bi-sine functions [29].

6.6.5. *Tracking of structures from an image sequence*

Participants: Yann Lepoittevin, Isabelle Herlin, Dominique Béréziat [UPMC].

The research concerns an approach to estimate velocity on an image sequence and simultaneously segment and track a given structure. It relies on the underlying dynamics' equations of the studied physical system. A data assimilation method is designed to solve evolution equations of image brightness, those of motion's dynamics, and those of distance map modeling the tracked structures. The method is applied on meteorological satellite data, in order to track tropical clouds on image sequences and estimate their motion.

Part of research was concerned on the numerical schemes applied for advecting the distance map and designing its adjoint.

6.7. Minimax filtering

In minimax filtering for state estimation, the initial state error, the model error and the observational errors are classically supposed to belong to one joint ellipsoid. In this case, it is only assumed that the errors, stochastic or deterministic, are bounded. For each assimilation experiment, the filter computes an ellipsoid where one will find at least all states compatible with observations and errors description. The state estimate is taken as the center of the ellipsoid. No assumption on the actual distribution of the errors in needed and the state estimate minimizes the worst-case error, which makes the filter robust.

6.7.1. A posteriori minimax motion estimation

Participants: Sergiy Zhuk [IBM Research, Ireland], Isabelle Herlin, Olexander Nakonechnyi [Taras Shevchenko National University of Kyiv], Jason Frank [CWI, the Netherlands].

Data assimilation algorithms based on the 4D-Var formulation look for the so-called conditional mode estimate. The latter maximizes the conditional probability density function, provided the initial condition, model error and observation noise are realizations of independent Gaussian random variables. However this Gaussian assumption is often not satisfied for geophysical flows. Moreover, the estimation error of the conditional mode estimate is not a first-hand result of these methods. The issues above can be addressed by means of the Minimax State Estimation (MSE) approach. It allows to filter out any random (with bounded correlation operator) or deterministic (with bounded energy) noise and assess the worst-case estimation error.

The iterative MSE algorithm was developed for the problem of optical flow estimation from a sequence of 2D images. The main idea of the algorithm is to use the "bi-linear" structure of the Navier-Stokes equations and optical flow constraint in order to iteratively estimate the velocity. The algorithm consists of the following parts:

1) we construct pseudo-observations $\widehat{I}$ as the estimate of the image brightness function $I(x,y,t)$ solving the optical flow constraint such that $\widehat{I}$ fits (in the sense of least-squares) the observed sequence of images. To do so, we set the velocity field in the optical flow constraint to be the current minimax estimate of the velocity field $\mathbf{w}$, obtained at the previous iteration of the algorithm, and construct the minimax estimate $\widehat{I}$ of the solution of the resulting linear advection equation using the observed image sequence as discrete measurements of the brightness function;

2) we plug the estimate of the image gradient, obtained out of pseudo-observations $\widehat{I}$ in 1), into the optical flow constraint and the current minimax estimate $\mathbf{w}$ of the velocity field into the non linear part of Navier-Stokes equations so that we end up with a system of linear PDEs, which represents an extended state equation: it contains a linear parabolic equation for the velocity field and linear advection equation for the image brightness function; we construct the minimax estimate of the velocity field from the extended state equation using again the observed image sequence as discrete measurements of the brightness function;

3) we use the minimax estimate of the velocity field obtained in 2) in order to start 1) again.

Point 1) has been implemented and tested. As Point 2) is currently under development, it is replaced by one of our motion estimation method in order to be plugged in Point 3).

6.8. Fire application

6.8.1. Model evaluation for fire propagation

Participants: Vivien Mallet, Jean-Baptiste Fillipi [CNRS], Bahaa Nader [University of Corsica].

In the field of forest fires risk management, important challenges exist in terms of people and goods preservation. Answering to strong needs from different actors (firefighters, foresters), researchers focus their efforts to develop operational decision support system tools that may forecast wildfire behavior. This requires the evaluation of models performance, but currently, simulation errors are not sufficiently qualified and quantified. As the main objective is to realize a *decision support system*, it is required to establish robust forecast evaluations. In the context of the ANR project IDEA, the evaluation of model simulations has led to the definition and implementation of a series of forecast scores. The merits and shortcomings of the scores were evaluated on synthetic cases. This demonstrated the efficiency of scores that take into account the fire dynamics, where some classical scores may fail. This was also found on real fires, using field observations.

In addition, we consider that the proper evaluation of a model requires to apply it to a large number of fires – instead of carrying out a fine tuning on just one fire. We implemented a software to simulate a large number of fires (from the Prométhée database, http://www.promethee.com/) with the simulation model ForeFire (CNRS/University of Corsica) and evaluate the results with error measures. One simulation requires mainly the following data: the ignition point, the ground elevation, the vegetation cover and the wind field. See illustration on Fig. 6 .

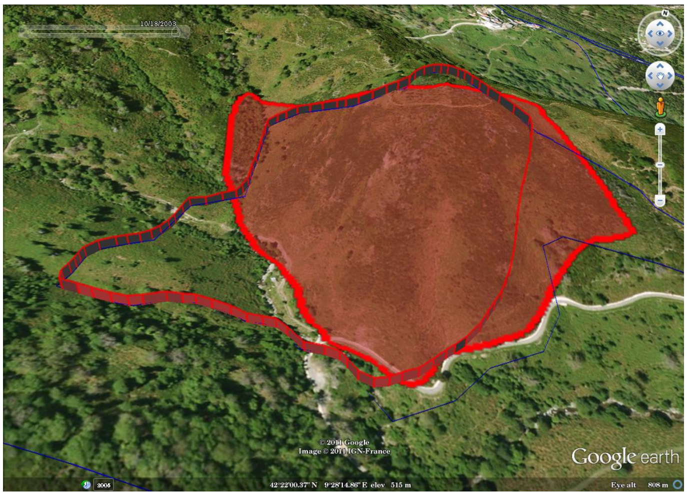


Figure 6. Fire simulation (using ForeFire) in red elevated contour, and observation (from Prométhée) of the burned area in filled red contour, for a 2003 fire near San-Giovanni-di-Moriani (Corsica).

CORTEX Project-Team

6. New Results

6.1. Spiking neurons

Participants: Hana Belmabrouk, Dominique Martinez, Thierry Viéville, Thomas Voegtlin.

6.1.1. Mathematical modeling

In order to understand the dynamics of spiking neural networks under the influence of a modified synaptic dynamics of single neurons, we study the effect of tonic inhibition on the population activity in spiking neural networks. The aim is to derive mathematical relations of the population activity and some statistics estimated numerically from the simulation of networks [4], [8].

6.1.2. Biophysical modeling

Our understanding of the computations that take place in the human brain is limited by the extreme complexity of the cortex, and by the difficulty of experimentally recording neural activities, for practical and ethical reasons. The Human Genome Project was preceded by the sequencing of smaller but complete genomes. Similarly, it is likely that future breakthroughs in neuroscience will result from the study of smaller but complete nervous systems, such as the insect brain or the rat olfactory bulb. These relatively small nervous systems exhibit general properties that are also present in humans, such as neural synchronization and network oscillations. Our goal is therefore to understand the role of these phenomena by combining biophysical modelling and experimental recordings, before we can apply this knowledge to humans. In the last year, we have studied new aspects of our models of the insect olfactory system [7], [14].

6.1.3. Using event-based metric for event-based neural network weight adjustment

The problem of adjusting the parameters of an event-based network model is addressed here at the programmatic level. Considering temporal processing, the goal is to adjust the network units weights so that the outcoming events correspond to what is desired. The work of [18] proposes, in the deterministic and discrete case, a way to adapt usual alignment metrics in order to derive suitable adjustment rules. At the numerical level, the stability and unbiasness of the method is verified.

The key point, here, is the non-learnability of even-based , since it is proved that this problem is NP-complete, when considering the estimation of both weights in the general case, except for exact simulation. We show that we can “elude” this caveat and propose an alternate efficient estimation mechanism, inspired by alignment metrics used in spike train analysis, thus providing a complement of other estimation approaches, beyond usual convolution metric. At last, the proposed mollification *is* a series of convolution metric, but that converges towards the expected alignment metric.

6.1.4. Predictive learning

In collaboration with Sander Bohte (CWI, Netherlands) and Nicolas Fourcaud-Trocme (CNRS, Lyon), we are developing a model of predictive learning using oscillations in a population of spiking neurons. The model is based on previous work performed in the Cortex group. Our previous model suggested a possible role for neuronal synchronization in unsupervised, predictive-type learning. However, that model was not compatible with sustained oscillations observed in biological networks. We are extending our initial approach in order to allow the network to learn during a stable, steady-state oscillatory regime. This extension involves using type-2 neurons and two distinct types of inhibition.

6.2. Dynamic Neural Fields

Participants: Frédéric Alexandre, Yann Boniface, Laurent Bougrain, Georgios Detorakis, Hervé Frezza-Buet, Bernard Girau, Axel Hutt, Mathieu Lefort, Nicolas Rougier, Wahiba Taouali.

The work reported this year represents both extensions of previous works and new results linked to the notion of neural population, considered at (i) a formal level (theoretical studies of neural fields), (ii) a numerical level (study of functioning and learning rules) and (iii) a more embodied one (implementations of specific functions).

6.2.1. Formal Level

To study the effect of external stimuli on nonlinear neural population dynamics involving constant delays, the work aims to apply the center manifold theorem and derive expressions of the time-dependent centre manifold. It is observed that additive noise and external quasi-periodic driving change the stability of neural populations dependent on the delay [9], [10].

6.2.2. Numerical Level

At the numerical level, specific developments were carried out to assess our software platform, to master functioning rules and to study the performances of new learning rules:

- Adaptation of the BCM rule to multi-modality by adapting the dynamics of the threshold by the use of a feed-back signal generated by a neural field map [1], [26]
- We investigate the formation and maintenance of ordered topographic maps in the primary somatosensory cortex as well as the reorganization of representations after sensory deprivation or cortical lesion. We consider both the critical period (postnatal) where representations are shaped and the post-critical period where representations are maintained and possibly reorganized. We hypothesize that feed-forward thalamocortical connections are an adequate site of plasticity while cortico-cortical connections are believed to drive a competitive mechanism that is critical for learning. We model a small skin patch located on the distal phalangeal surface of a digit as a set of 256 Merkel ending complexes (MEC) that feed a computational model of the primary somatosensory cortex (area 3b). This model is a two-dimensional neural field where spatially localized solutions (a.k.a. bumps) drive cortical plasticity through a Hebbian-like learning rule. Simulations explain the initial formation of ordered representations following repetitive and random stimulations of the skin patch. Skin lesions as well as cortical lesions are also studied and results confirm the possibility to reorganize representations using the same learning rule and depending on the type of the lesion. For severe lesions, the model suggests that cortico-cortical connections may play an important role in complete recovery [6].

6.2.3. Embodied Level

6.2.3.1. Motion detection

We develop bio-inspired neural architectures to extract and segment the direction and speed components of the optical flow from sequences of images. Following this line, we have built additional models to code and distinguish different visual sequences. The structure of these models takes inspiration from the course of visual movement processing in the human brain, such as in area MT (middle temporal) that detects patterns of movement, or area FBA where neurons have been found to be sensitive to single spatio-temporal patterns. This work has been extended to complex movements: to fight, to wave, to clap, using real-world video databases [5].

6.2.3.2. Anticipatory mechanisms in neural fields

We have defined first models of neural fields that include anticipatory mechanisms through the integration of spatiotemporal representations into the lateral interactions of a dynamic neural field. In [20], the case of multiple anticipated trajectories is studied.

6.2.3.3. Action selection

Within the context of enaction and a global approach to perception, we focused on the characteristics of neural computation necessary to understand the relationship between structures in the brain and their functions. We first considered computational problems related to the discretization of differential equations that govern the studied systems and the synchronous and asynchronous evaluation schemes. Then, we investigated a basic functional level : the transformation of spatial sensory representations into temporal motor actions within the visual-motor system. We focused on the visual flow from the retina to the superior colliculus to propose a minimalist model of automatic encoding of saccades to visual targets. This model, based on simple local rules (CNFT and logarithmic projection) in a homogeneous population and using a sequential processing, reproduces and explains several results of biological experiments. It is then considered as a robust and efficient basic model. Finally, we investigated a more general functional level by proposing a computational model of the basal ganglia motor loop. This model integrates sensory, motor and motivational flows to perform a global decision based on local assessments. We implemented an adaptive process for action selection and context encoding through an innovative mechanism that allows to form the basic circuit for other cortico-basal loops. This mechanism allows to create internal representations according to the enactive approach that opposes the computer metaphor of the brain. Both models have interesting dynamics to study from whether a biological point of view or a computational numerical one [2], [12].

6.3. Higher level functions

Participants: Frédéric Alexandre, Laurent Bougrain, Octave Boussaton, Axel Hutt, Maxime Rio, Carolina Saavedra, Christian Weber.

Our activities concerned information analysis and interpretation and the design of numerical distributed and adaptive algorithms in interaction with biology and medical science. To better understand cortical signals, we choose a top-down approach for which data analysis techniques extract properties of underlying neural activity. To this end several unsupervised methods and supervised methods are investigated and integrated to extract features in measured brain signals. More specifically, we worked on Brain Computer Interfaces (BCI).

6.3.1. Using Neuronal States for Transcribing Cortical Activity into Muscular Effort

We studied the relations between the activity of corticomotoneuronal (CM) cells and the forces exerted by fingers. The activity of CM cells, located in the primary motor cortex is recorded in the thumb and index fingers area of a monkey. The activity of the fingers is recorded as they press two levers. The main idea of this work is to establish and use a collection of neuronal states. At any time, the neuronal state is defined by the firing rates of the recorded neurons. We assume that any such neuronal state is related to a typical variation (or absence of variation) in the muscular effort. Our forecasting model uses a linear combination of the firing rates, some synchrony information between spike trains and averaged variations of the positions of the levers [17].

6.3.2. From the decoding of cortical activities to the control of a JACO robotic arm: a whole processing chain

We realized a complete processing chain for decoding intracranial data recorded in the cortex of a monkey and replicates the associated movements on a JACO robotic arm by Kinova. We developed specific modules inside the OpenViBE platform in order to build a Brain-Machine Interface able to read the data, compute the position of the robotic finger and send this position to the robotic arm. More precisely, two client/server protocols have been tested to transfer the finger positions: VRPN and a light protocol based on TCP/IP sockets. According to the requested finger position, the server calls the associated functions of an API by Kinova to move the fingers properly. Finally, we monitor the gap between the requested and actual fingers positions. This chain can be generalized to any movement of the arm or wrist [22].

6.3.3. Wavelet-based Semblance for P300 Single-trial Detection

Electroencephalographic signals are usually contaminated by noise and artifacts making difficult to detect Event-Related Potential (ERP), specially in single trials. Wavelet denoising has been successfully applied to ERP detection, but usually works using channels information independently. This paper presents a new adaptive approach to denoise signals taking into account channels correlation in the wavelet domain. Moreover, we combined phase and amplitude information in the wavelet domain to automatically select a temporal window which increases class separability. Results on a classic Brain-Computer Interface application to spell characters using P300 detection show that our algorithm has a better accuracy with respect to the VisuShrink wavelet technique and XDAWN algorithm among 22 healthy subjects, and a better regularity than XDAWN [21].

6.3.4. Filter for P300 detection

According to recent literature, the most appropriate preprocessing to improve P300 detection is still unknown or at least there is no consensus about it. Research papers refer to different low-pass filters, high-pass filters, baseline, subsampling or feature selection. Using a database with 23 healthy subjects we compared the effect on the letter accuracy (single-trial detection) provided by a linear support vector machine of a high-pass filter with cutoff frequencies from 0.1 to 1 Hz and a low-pass filter with cutoff frequencies from 8 to 60 Hz. According to this study, the best combination is for a band-pass filter of 0.1 to 15 Hz [16].

6.3.5. Processing Stages of Visual Stimuli and Event-Related Potentials

Event-evoked potentials (ERP) in electroencephalograms reflect various visual processing stages according to their latencies and locations. Thus, ERP components such as the N100, N170 and the N200 which appears 100, 170 and 200 ms after the onset of a visual stimulus correspond respectively to a selective attention, the processing of color, shape and rotation (e.g. processing of human faces) and a degree of attention [24].

6.3.6. Exploring the role of the thalamus in visuomotor tasks implicating non-standard ganglion cells

Non-standard ganglion cells in the retina have specific loci of projection in the visuomotor systems and particularly in the thalamus and the superior colliculus. In the thalamus, they feed the konio pathway of the LGN. Exploring the specificities of that pathway, we discovered it could be associated to the matrix system of thalamo-cortical projections, known to allow for diffuse patterns of connectivity and to play a major role in the synchronization of cortical regions by the thalamus.

An early model [23] led to the design of the corresponding information flows in the thalamo-cortical system, that we are now expanding, in the framework of the Keops project § 7.2 , to be applied to real visuomotor tasks.

6.3.7. Formalization of input/output retinal transformation regarding non-standard gangion cells behavior

We propose to implement the computational principles raised by the study on the K-cells of the retina using a variational specification of the visual front-end, with an important consequence: In such a framework, the GC are not to be considered individually, but as a network, yielding a mesoscopic view of the retinal process.

Given natural image sequences, fast event-detection properties appears to be exhibited by the mesoscopic collective non-standard behavior of a subclass of the so-called dorsal and ventral konio-cells (K-cells) that correspond to specific retinal output.

We consider this visual event detection mechanism to be based on image segmentation and specific natural statistical recognition, including temporal pattern recognition, yielding fast region categorization. We discuss how such sophisticated functionalities could be implemented in the biological tissues as a unique generic two-layered non-linear filtering mechanism with feedback. We use computer vision methods to propose an effective link between the observed functions and their possible implementation in the retinal network.

The available computational architecture is a two-layers network with non-separable local spatio-temporal convolution as input, and recurrent connections performing non-linear diffusion before prototype based visual event detection.
The numerical robustness of the proposed model has been experimentaly checked on real natural images. Finally, model predictions to be verified at the biological level are discussed [25].

6.4. Embodied and embedded systems

Participants: Yann Boniface, Hervé Frezza-Buet, Bernard Girau, Mathieu Lefort.

6.4.1. InterCell

Our research in the field of dedicated architectures and connectionist parallelism mostly focuses on embedded systems (*cf.* §3.5). Nevertheless we are also involved in a project that considers coarse-grain parallel machines as implementation devices. The core idea of this InterCell project (*cf.*http://intercell.metz.supelec.fr) is to map fine grain computation (cells) to the actual structure of PC clusters. The latter rather fit coarse grain processing, using relatively few packed communication, which a priori contradicts neural computing. Another fundamental feature of the InterCell project is to promote interaction between the parallel process and the external world. Both features, cellular computing and interaction, allow to consider the use of neural architectures on the cluster on-line, for the control of situated systems, as robots.

6.4.2. Hardware implementations of neural models

In the field of dedicated embeddable neural implementations, we use our expertise in both neural networks and FPGAs so as to propose efficient implementations of applied neural networks on FPGAs, as well as to define hardware-friendly neural models.

- We currently intend to minimize the topological constraints of FPGA-embedded spiking neural fields using reduced neighborhoods but randomly propagating spikes. A preliminary result has been obtained so as to implement massively distributed pseudo-random number generators based on cellular automata that use minimal areas though they produce random streams that pass most randomness tests [19]. These results have also been applied to cellular aurtomata using randomnes in their transition rules [13].
- Researchers have proposed the concept of Central Pattern Generators (CPGs) as a neural mechanism for generating an efficient control strategy for legged robots based on biological locomotion principles. We have developed a reconfigurable hardware implementation of a CPG-based controller which is able to generate several gaits for quadruped and hexapod robots [3].

6.4.3. Towards brain-inspired hardware

Our activities on dedicated architectures have strongly evolved in the last years. We now focus on the definition of brain-inspired hardware-adapted frameworks of neural computation. Our current works aim at defining hardware-compatible protocols to assemble various perception-action modalities that are implemented and associated by different bio-inspired neural maps.

6.4.3.1. Multimodal learning through joint dynamic neural fields

This work relates to the development of a coherent multimodal learning for a system with multiple sensory inputs. We have modified the BCM synaptic rule, a local learning rule, to obtain the self organization of our neuronal inputs maps and we use a CNFT based competition to drive the BCM rule. In practice, we introduce a feedback modulation of the learning rule, representing multimodal constraints of the environment, and we introduce an unler-arning term in the BCM equation to solve the problem of the different temporalities between the raise of the activity within modal maps and the multimodal learning of the organization of the maps [1], [26].

6.4.3.2. Randomly spiking dynamic neural fields

We have defined a new kind of spiking neural field that is able to use only local links while transmitting spikes through the map by succesive random propagations. Such a model is able to be mapped onto FPGAs, while maintaining most properties of neural fields. Early results will be soon published.

DEMAR Project-Team

6. New Results

6.1. Modelling and Identification

6.1.1. Subject-specific Center of Mass estimation in human subjects

Participants: Alejandro González, Mitsuhiro Hayashibe, Philippe Fraisse.

Center of mass position (CoM) in humanoid robots can be used to generate a joint trajectory suitable for walking and standing. Oscillations of the CoM while maintaining a standing posture have been observed in older patients. These oscillations are thought to occur due a change in the subject's balancing strategy. This is why the motion of the CoM in generally considered as good metric to be used while diagnosing pathologies which affect gait. With this in mind we propose the use of the statically equivalent serial chain (SESC) to provide a subject specific estimate of CoM position. The state of the art techniques for CoM estimation in humans involve the use of expensive equipment in a laboratory setting, making it difficult to use as a clinical tool or inside the home environment. Current work for diagnosing a subject's balance moves away from this, focusing on wearable and minimally invasive sensors that obtain information during the subject's daily activities. We propose the use of widely available sensors like the Kinect camera, for tracking the subject's movements, and the Wii balance board during the calibration of the SESC.

a) b)

Figure 1. Center of mass estimation can be done using the statically equivalent serial chain. We assume a 9 rigid-link model (a) with spherical joints, capable of three dimensional movements. The SESC estimation can be performed in real time and is driven by joint angular measurements (b).

We have focused in improving the reliability of the identification during a study of the CoM trajectory in the sagital plane [14]. We studies small value for the condition number of the used data as well as of the parameter relative standard deviation ($\sigma_{\hat{\mathbf{R}}_r}\%$) are useful to determine the validity of the estimate. Subsequent works have extended the human model to three-dimensional motion (Fig. 1 .a). In order to observe the tracking of the center of mass, we have developed a 3D visualization tool which represents the subject's skeleton and SESC in real time (Fig. 1 .b) [16]; it is also possible to observe the CoM history. Finally we are now comparing the performance of the inexpensive Kinect sensor and the traditional video based motion capture Vicon system, with reasonably good results [15]. Current work is also focused in improving the speed of the identification phase by simplifying the assumed human model and using physical constraints to reduce the complexity of the SESC. Also the development of a simple to follow identification protocol which takes into account multiple supporting surfaces is desired.

6.1.2. Recursive estimation of SESC parameters for human Center of Mass estimation

Participants: Alejandro González, Mitsuhiro Hayashibe, Philippe Fraisse.

A human's center of mass (CoM) trajectory is useful to evaluate the dynamic stability during daily life activities such as walking and standing up. To estimate the subject-specific CoM position in the home environment, we make use of a statically equivalent serial chain (SESC) developed with a portable mesurement system. In order to adapt to the subject's physical capacities we implement a constrained Kalman filter to achieve an online parameter estimation of the SESC parameters. By accounting for the human body bilateral symmetry we hope to reduce the identification time. This results in constrainig SESC parameters to be consistent with the human skeletal model used. The Kinect camera is used as a markerless motion capture system for measuring limb orientations while the Wii board is used to measure the subject's center of pressure (CoP) during the identification phase. For his experiment CoP measurements and Kinect data were recorded for five able-bodied subjects. The data was then given to the proposed recursive algorithm to identify the parameters of the SESC online (Fig.2). This method of online identification allows the subject or the therapist to know the quality of the on-going CoM identification while giving postures; by doing this the time needed to perform the identification can be reduced. A cross-validation was performed to verify the identification performance.

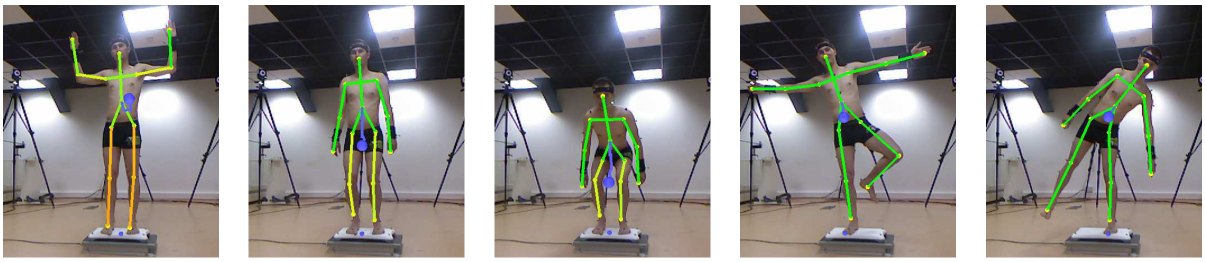


Figure 2. Online statically equivalent serial chain (SESC) parameter estimation. The length of each of the SESC's link is updated when a static pose has been found and can be observed in real time. The color of the skeleton can be updated as a cue to the subject and/or therapist.

6.1.3. FES-Induced Torque Prediction with Evoked EMG Synthesized by Recurrent Neural Network

Participants: Zhan Li, Mitsuhiro Hayashibe, David Guiraud.

A NARX-type recurrent neural network (NARX-RNN) model is proposed for identification and prediction of FES-induced muscular dynamics with eEMG. Such NARX-RNN model is with a novel architecture for prediction, with robust prediction performance. To make fast convergence for identification of such NARXRNN, directly-learning pattern is exploited during the learning phase. Due to difficulty of choosing a proper forgetting factor of Kalman filter for predicting time-variant torque with eEMG, such NARX-RNN may be considered to be a better alternative as torque predictor. Data gathered from two SCI patients is used to evaluate the proposed NARX-RNN model. The NARX-RNN model shows promising estimation and prediction performance only based on eEMG [23].

6.1.4. Inverse Estimation of Muscle Activations with Weights Optimization

Participants: Zhan Li, Mitsuhiro Hayashibe, David Guiraud.

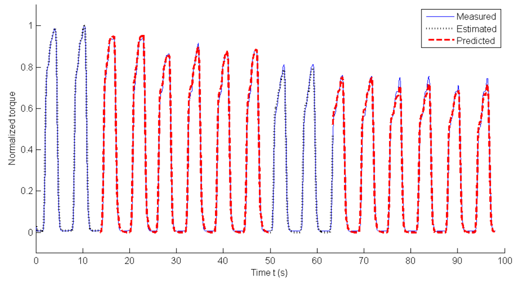


Figure 3. Performance of NARX-RNN model with eEMG for prediction with identification on periodically

Inverse estimation of activations of muscle groups at human lower leg in random movement condition is investigated with merely the ankle joint torque used. Optimization technique for the relationship between muscle activations and torque is exploited. Such optimization is able to rebuilt the relationship between muscle activations and torque inversely based on experimental data obtained from five healthy subjects, and the optimal weight matrix can indicate each muscle's contribution for producing the torque. Further cross-validation on prediction of muscle activations with joint torque with optimal weights shows such approach may possess promising performance [22].

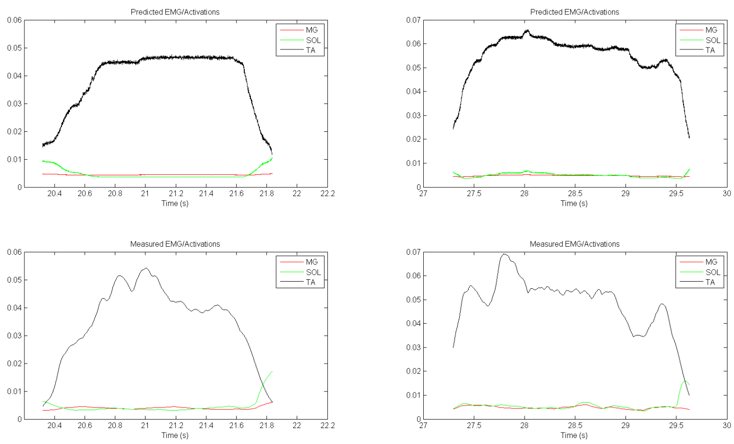


Figure 4. Prediction of muscle activations

6.1.5. 3D Volumetric Muscle Modeling For Real-time Deformation Analysis With FEM

Participants: Yacine Berranen, Mitsuhiro Hayashibe, Benjamin Gilles, David Guiraud.

Computer simulators are promising numerical tools to study muscle volumetric deformations but most models are facing very long computation time and thus are based on simplified Hill model versions. The purpose of this study is to develop a real-time three-dimensional biomechanical model of fusiform muscle based on modified Hill model for the active stress which is controlled from EMG recordings. Finite element model is used to estimate the passive behavior of the muscle and tendons during contraction. We show that this 3D model implementation is very cost effective with respect to the computation time and the simulation gives good results compared to real measured data. Thus, this effective implementation will allow implementing much more complex and realistic models, with moderate computation time.

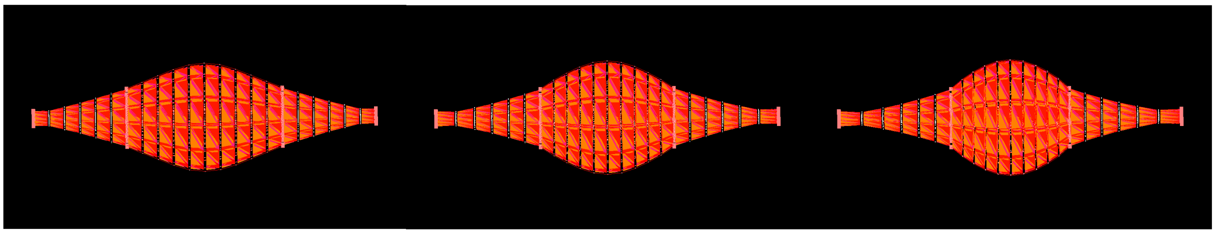


Figure 5. Real-time Simulation of Volumetric Muscle Deformation.

6.1.6. Principal Geodesic Dynamics

Participants: Maxime Tournier, Lionel Reveret.

This paper presents a new integration of a data-driven approach using dimension reduction and a physically-based simulation for real-time character animation. We exploit Lie group statistical analysis techniques (Principal Geodesic Analysis, PGA) to approximate the pose manifold of a motion capture sequence by a reduced set of pose geodesics. We integrate this kinematic parametrization into a physically-based animation approach of virtual characters, by using the PGA-reduced parametrization directly as generalized coordinates of a Lagrangian formulation of mechanics. In order to achieve real-time without sacrificing stability, we derive an explicit time integrator by approximating existing variational integrators. Finally, we test our approach in task-space motion control. By formulating both physical simulation and inverse kinematics time stepping schemes as two quadratic programs, we propose a features-based control algorithm that interpolates between the two metrics. This allows for an intuitive trade-off between realistic physical simulation and controllable kinematic manipulation.

6.1.7. An improved kinematic model of the spine for three-dimensional motion analysis in the Vicon system

Participants: Pawel Maciejasz, Wieslaw Chwala (University School of Physical Education, Krakow), Miroslawa Dlugosz, Daria Panek, Witold Alda (AGH University of Science and Technology, Krakow).

The mechanism of creation and pathomechanics of lateral spinal deformation is still not fully explained. Modern medical imaging techniques give scientists possibility to understand some aspects, but vast majority of those techniques is based on static trials. A motion capture system belongs to techniques which enable visualization of a spine during dynamic trials; however, due to lack of appropriate computational model, it is unsuitable for scoliosis imaging.

A few years ago a kinematic model of the spine has been proposed to be used with Vicon Motion Capture System (Master thesis of P. Maciejasz). This model was based on Bézier curves and allowed for much more precise investigation of spinal kinematics during dynamic trials as compared with other computational models. However, it did not allowed to restrict only selected movements for particular segments of the spine (e.g. axial rotation for lumbar spine). The aim of the current work is to improve the proposed model in order to be able to restrict selected movements according to the knowledge concerning spinal anatomy and spinal range of motion. The new kinematic model of the spine was written in BodyBuilder for Biomechanics Language.

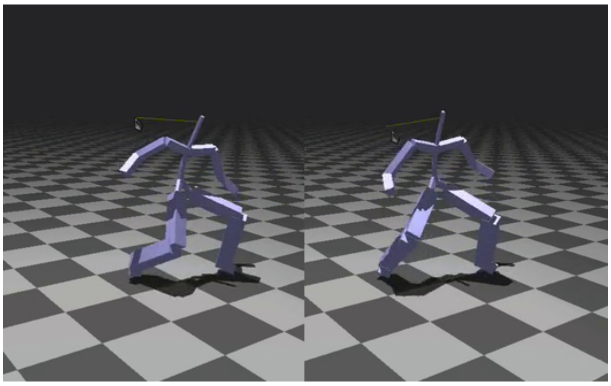


Figure 6. Real-time user interaction with a virtual character maintaining balance, animated using our approach: based on a break-dance motion capture sequence, our dimension reduction process allows one to compute both dynamics and task-space control in a low-dimensional, data-driven subspace.

For the purpose of visualization also an accurate graphical representation of each vertebra (polygon mesh) was computed and adapted to be compatible with the kinematic model. Using a new version of the model it is possible to perform precise analysis of movement of all vertebrae during such dynamic activities as e.g. gait and forward or lateral bending, as well as to present the results not only on the charts, but also as a 3D animation of movements of a realistically looking spine [12].

6.1.8. Methodology of automated detection and classification of action potentials in nerve fiber based on multichannel recordings

Participants: Thomas Guiho, Pawel Maciejasz, David Guiraud.

For some neuroprosthetic applications it would be beneficial to be able to automatically detect when particular nerve fibers (axons) are in "firing" (i.e. when an action potential is propagating along them). Due to limitations of currently available recording electrodes on one side, and the attempt to be as little invasive as possible, at the moment in practical application it is not possible to record signals coming from a single nerve fiber. In signal recorded using typical electrodes placed close to nerve fibers, action potentials coming from various nerve fibers, as well as noise coming from outside of the nerve, may be detected. One of the possibilities allowing to distinguish action potentials coming from various nerve fibers in such a case is to record signals at a few places along the nerve and compare them.

We have proposed an algorithm that allows to automatically detect and classify evoked action potentials in a simple earthworm model. The signals were recorded concurrently at 2 places along the giant nerve fibers. In the first step the algorithm tries to identify only does recordings in which action potentials generated by various nerve fibers can be easily distinguished. Afterwards, the most significant features (such as amplitude, duration, propagation velocity, etc.) to distinguish between different populations of fibers are identified. Finally, the action potentials in all signals are identified and classified using the features determined in the previous step.

The proposed method was implemented using MATLAB software and tested on the file containing almost 200 signals record in response to various stimuli. The same data ware later inspected manually and the action potentials were manually classified. More than 99% of action potentials were classified to the same nerve fiber when performing automatic and manual classification.

6.2. Function control and synthesis

6.2.1. FES assisted sitting pivot transfer

Participants: Jovana Jovic, Christine Azevedo Coste, Philippe Fraisse, Sebastien Lengagne, Charles Fattal.

Transferring from a wheelchair to a treatment table, bed, tub/shower bench, toilet seat, car seat and vice versa represent typical Sitting Pivot Transfer (SPT) realized by individuals with Spinal Cord Injury (SCI). Individuals with SCI, perform this postural task around fifteen times a day using upper extremities. In the chronic stage after SCI, soft tissue structures are exposed to overuse in activities of daily living, such as, transfer task in which the shoulder becomes a weight-bearing joint. Therefore, the risk of shoulder pain and musculoskeletal disorders is higher in persons with paraplegia compared with an able-bodied population. A lot of scientific effort has been focused on experimental studies in which the kinetic and the kinematic of the SPT movement have been analyzed. To our best knowledge, the scientists have focused their attention only on the performance of SPT; the influence of Functional Electrical Stimulation (FES) on SPT maneuver has not been investigated so far.

Therefore, we investigate the influence of FES on SPT motion of a paraplegic person. First, we develop dynamic optimization method in order to predict SPT motion of an able-bodied subject. This approach have been validated by comparing computed SPT trajectories with the ones measured during the experiment with an able-bodied subject (see Fig. 7). After validating our method, we used the optimization tool for analyzing the influence of FES on SPT maneuver in paraplegic persons. Our results suggest that FES can decrease arm participations during the transfer motion of a paraplegic person. [6], [21].

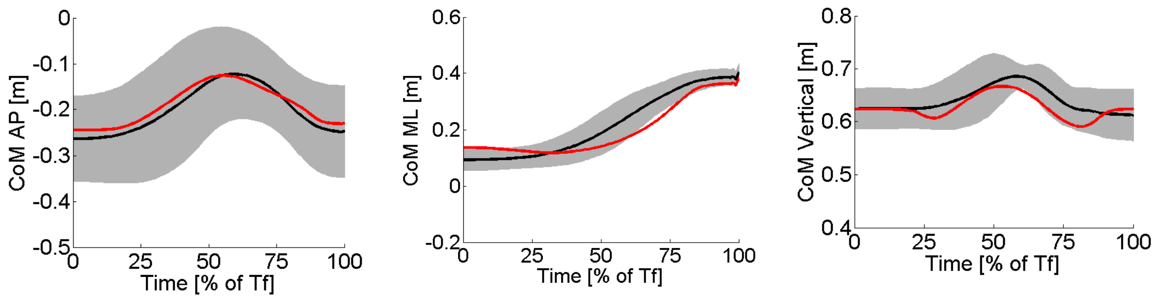


Figure 7. Computed (red line) CoM position, mean value of CoM positions estimated from experimental data (black line) and its plus/minus standard deviation (gray line) in AP, ML, and vertical direction in able-bodied subject.

6.2.2. FES assisted prolonged standing

Participants: Jovana Jovic, Philippe Fraisse, Christine Azevedo Coste, Charles Fattal.

Prolonged immobilization which occurs after spinal cord injury results in many physiological problems. Standing therapy can ameliorate many of those problems. The approaches proposed in the literature for restoration of standing in paraplegic population based on Functional Electrical Stimulation (FES) focuse on the control of each individual joint, i.e. joint space control. In those cases the balance of the postural system is not directly controlled. This could be problematic especially when only the lower limbs are controlled. During paraplegic's quiet standing two concurrent controllers are acting in parallel, the physiological system under control of Central Nervous System (CNS), and artificial FES system. Upper part of the paraplegic's body is under voluntary control, therefore artificial controllers should be designed in the way to take into account actions of the intact part of the body and to assist users in their task.

For human beings the Center of Mass (CoM) provides an indicator of stability and it is an essential parameter in human postural stability. By controlling CoM position in paraplegic person the voluntary motions under CNS control are taken into account. Therefore, in we propose a whole body controller based on control of the CoM position. The goal was to develop a simple balance controller which would, by means of FES, enable quiet standing of individuals suffering from SCI while taking into account the voluntary motion of the upper limbs. The controller should enable prolonged standing by simulating the behavior of an able-bodied subject during the standing task, i.e. by imposing posture switching and in that way allowing the stimulated muscles to relax. The proposed approach is based on a 10 DoF biomechanical model and Proportioan Integral (PI) controller (see Fig. 8). The validity of the approach is tested, in computer simulations, using human CoM trajectories estimated from experimental data and by applying perturbations in simulation during quiet standing in order to simulate voluntary upper body movements.The results show that proposed controller is able to track desired CoM position with sufficient precision and to maintain stability even in the presence of simulated movements of the upper body [20].

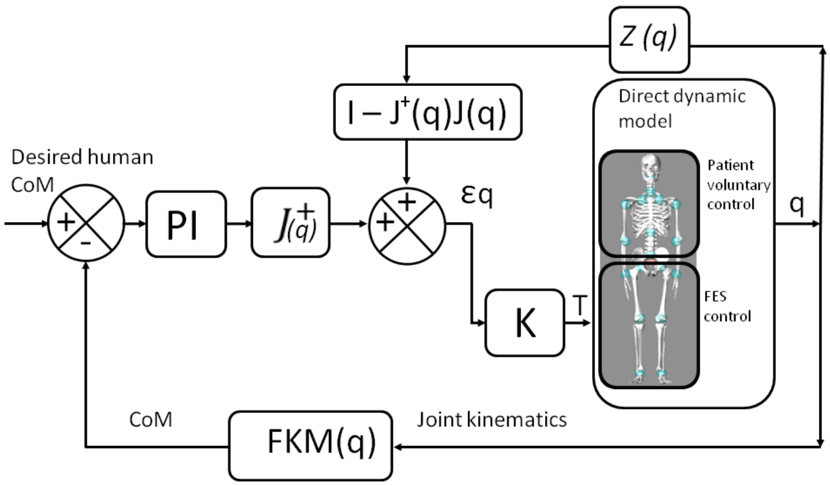


Figure 8. Block diagram of the proposed postural controller. Controller follows desired 3D CoM positions and controls the lower limbs by applying torque at ankle, knee and hip joints (10 DoF).

6.2.3. Bimanual reeducation assisted by FES in post-stroke patients

Participants: Nader Rouis, Christine Azevedo Coste, Philippe Fraisse, Isabelle Laffont, Denis Mottet.

This starting project will investigate the possibility to stimulate the deficient upper limb of a post-stroke patient in order to reproduce the movements observed on the valid upper limb or in order to achieve a bimanual task in cooperation with the valid upper limb. The aim is to improve the bimanual training tasks classically used in fictional rehabilitation. Both embedded sensors and kinect type systems will be investigated as possible ways to observe the valid upper limbs.

6.2.4. Freezing detection in Parkinson Disease patients

Participants: Maud Pasquier, Christine Azevedo Coste, Christian Geny, Bernard Espiau.

This work intends to apply the results of Maud Pasquier thesis about data segmentation and locomotion analysis to the detection as soon as possible of freezing episodes in Parkinson Disease (PD) patients. PD is a chronic degenerative disease of the central nervous system. One of the consequence is walking troubles and increased postural instability and falling risks. Freezing concerns at least half of PD patients, it is characterized by the transitory incapacity to make a step. It classically occurs at the gait initiation, turn around and passing doors. This freezing of gait (FOG) strongly impacts patient's mobility. As an example, in figure 9 , data recorded by an inertial sensor placed at the ankle is presented.

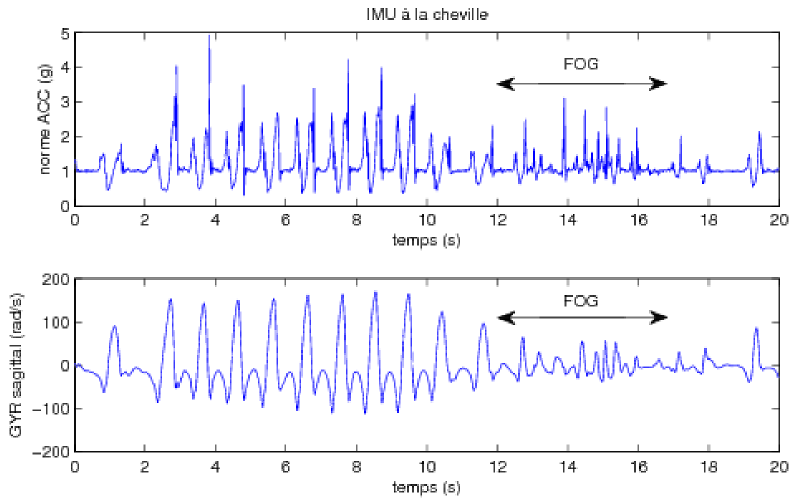


Figure 9. Signals recorded with an inertial sensor placed at the ankle during a walking trial. Top: acceleration norm, Bottom: sagittal plane rotation velocity

Assistive devices have been proposed, it has been shown that providing the patient with an auditory metronome or visual lines on the ground allows to reduce FOG occurrence but the effects are not maintained within the time. An approach would be to present this signals only when a freezing episode occurs and this implies to be able to detect it. Several authors have shown that FOG are in general associated to walking rhythm variability. Moore et al have used an accelerometer placed on lower limv in order to detect the presence of high frequencies. Indeed, the tremor pattern which can be observed during freezing is located at 3 et 8Hz, whereas normal locomotion is around 3Hz. These authors are able to detect a large part of the FOG but only those presenting high frequency patterns which is not systematic. Furthermore the detection delay (FFT) is very high and cannot be compatible with assistive device control constraints. We have proposed to observe stride properties in an online manner and compute a criterion which value informs about the FOG occurrence. The criterion is based on two variables: the stride length and the cadence. When stride length diminishes and cadence increases a FOG may be upcoming. 3 patients have been involved in this study and the criterion proposed has been shown to be as efficient than Moore's method in terms of number of detected FOG but the detection time is strongly improve with our method.

6.2.5. "Awake surgery" of slow-growing tumors and cortical excitability measured by EEG recordings.

Participants: François Bonnetblanc, Guillaume Herbet, Pom Charras, Mitsuhiro Hayashibe, David Guiraud, Hugues Duffau, Bénédicte Poulin-Charronnat.

Using direct electrical stimulation, real-time functional mapping of the brain can be used to perform resections of slow-growing infiltrative tumors in awake patients and to prevent the resection of essential areas near the tumor. To investigate interhemispheric imbalance following " awake surgeries" of slow-growing tumors we recorded EEG in a visuo- manual RT paradigm. Increase of cortical excitability within the ipsilesional hemisphere was signed by increased event related potentials (ERPs) amplitude for two patients. The cortical excitability in the lesioned hemisphere may be increased to maintain performances and cerebral plasticity.

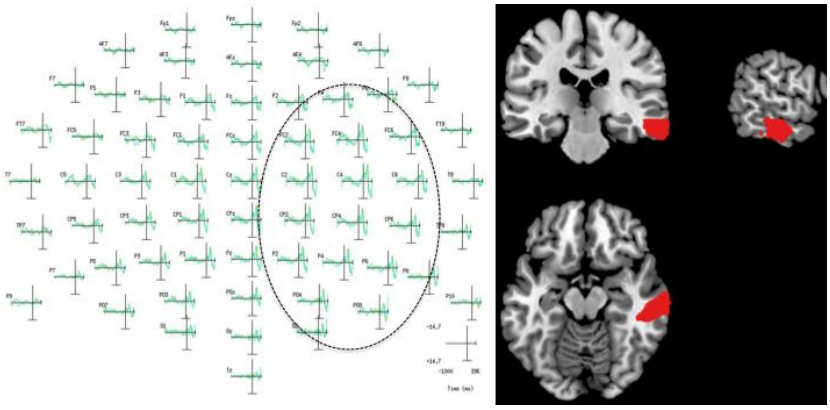


Figure 10. Left : Post-operative Event-Related Potentials (ERPs) for Patient 1. Increased ERPs amplitude can be seen in the right ipsilesional hemisphere (ellipse) in comparison to the contralesional hemisphere and homologous recording sites. The patient had to respond to visual stimuli occurring in the right or left hemifield with his right hand. The vertical line on each ERP indicates the occurrence of the visual go- signal Right : Lesion mapping for the same patient after the surgery.

6.3. Neuroprostheses

6.3.1. Distributed Measurement Unit for Closed-Loop Functional Electrical Stimulation: Prototype for Muscular Activity Detection

Participants: Guillaume Coppey, David Andreu, David Guiraud.

One way to face centralized Functional Electrical Stimulation (FES) architecture limitations is to distribute electronics close to electrodes. These Distributed Stimulation and Measurement Units (DSU and DMU) are interconnected by a network. Different DSU have been designed and prototyped. We started the design and prototyping of a DMU dedicated to ElectroMyoGramm (EMG) activity reading.

To validate both the digital architecture of the DMU and the digital processing it performs, we prototyped a DMU in charge of muscular activity detection. This DMU is able to detect a threshold crossing on an EMG input signal. The experimental setup is schematically represented on figure 11 , showing also the digital architecture of the prototyped DMU.

This DMU is able to accurately detect EMG activity after filtering and then processing a rolling average. Figure 12 shows intermediate signals: (a) is the absolute value of the filtered EMG signal, (b) is the rolling average, and (c) is the threshold used for activity detection.

This DMU prototype showed that digital processing chain dedicated to EMG activity detection can be embedded within a distributed measurement unit using a programmable logical device (FPGA), like we did for distributed stimulation unit. The embedded architecture of this unit is designed according to a Petri Net based methodology. This allows to exploit effective parallelism offered by FPGA devices, and to reach expected

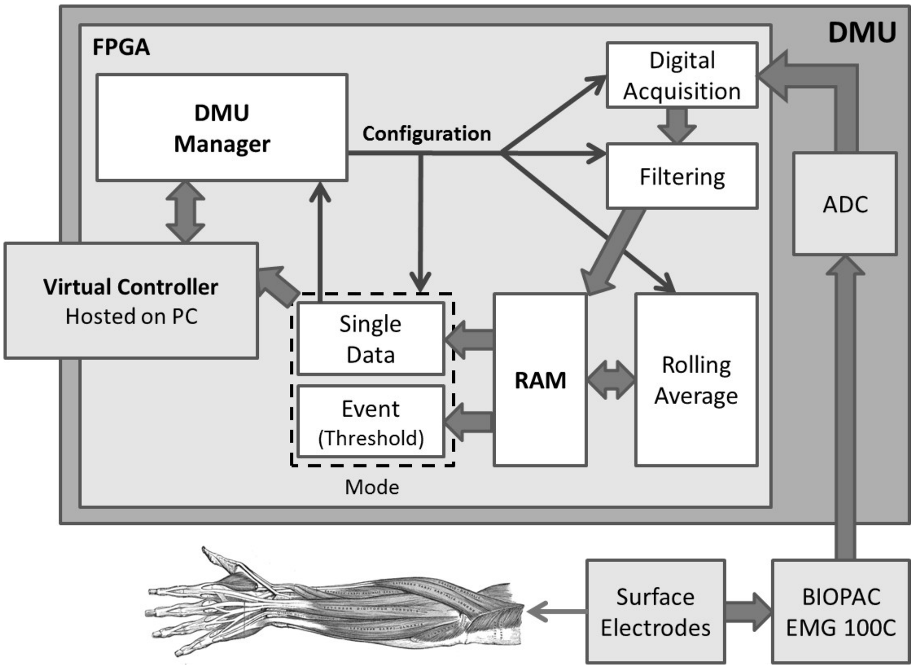


Figure 11. Experimental setup

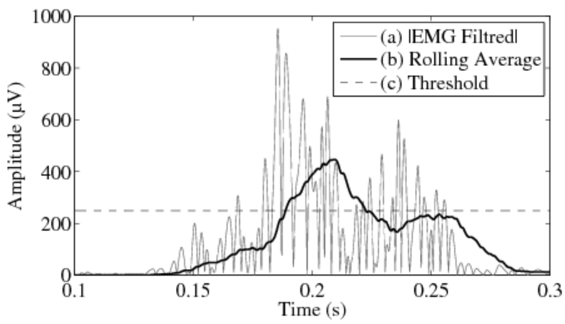


Figure 12. Intermediate signals for muscular activity detection on extensor digitorum communis

performances even at low frequency. The embedded processing chain is configurable and parameters can be adjusted, in order to optimize performance.

Future works will consist in adding the protocol stack to the digital architecture of the DMU, allowing integrating it within our distributed FES architecture. This will allow us to measure effective latencies and other performances from a closed-loop point of view. This work is necessary to ensure that such a distributed EMG activity detection is adequate with FES requirements. After that, we will investigate the trade-off between the global performances versus the implantable device constraints, like its size and power consumption.

6.3.2. Abstraction and composition for formal design of neuroprotheses

Participants: Hélène Leroux, David Andreu, Karen Godary.

In the framework of specification and implementation of complex digital systems on FPGA, we have developped an approach based on components whose behavior and composition are specified by generalized interpreted T-time Petri nets. One of the inherent difficulties for designer is to take into account, on the behavioral part, exceptions. This leads often to a complex modeling and is a source of human errors. Indeed, it is intricate to express all the possible situations (i.e. current state of model). We have defined a way to model exception handling by integrating the well-know concept of macroplace into the formalism. The analysability of the model and the efficiency of the implementation on FPGA (reactivity and surface, ie number of logic blocks) have been preserved. An example of macroplace is given in figure 13 ; it contains a sub-net (set of places of its refinement) from which exception handling is simply described by a dedicated output transition (transition t_e on fig. 13), whatever is the current state of the sub-net.

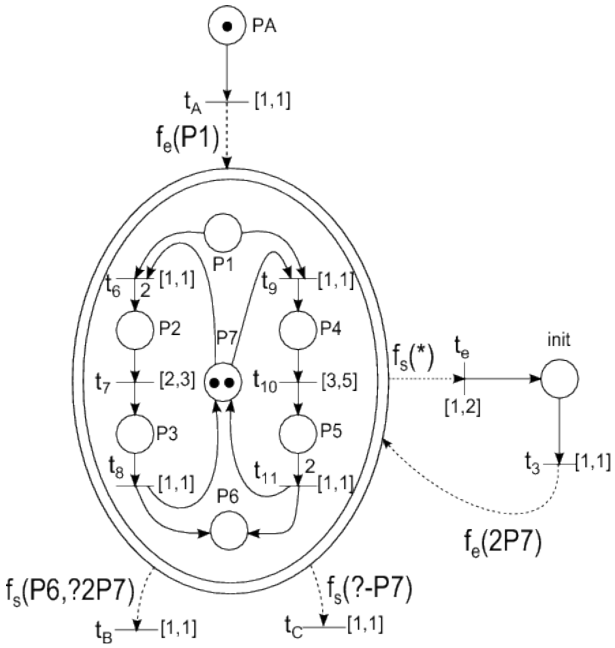


Figure 13. An example of macroplace and exception catching

6.3.3. Increasing stimulation selectivity using the Transversal Intrafascicular Multichannel Electrode (TIME)

Participants: Pawel Maciejasz, David Andreu, David Guiraud, Xavier Navarro, Jordi Badia (Universitat Autònoma de Barcelona), Winnie Jensen, Kristian Rauhe Harreby, Aritra Kundu, Bo Geng (Aalborg University), Thomas Stieglitz, Tim Boretius (University of Freiburg), Ken Yoshida (Purdue University Indianapolis).

The electrical stimulation of nerve fibres may allow to restore or augment some body functions lost due to disease or injury. However, in typical peripheral nerves there are thousands of nerve fibres innervating various organs. Therefore, it is necessary to develop interfaces and methods allowing for selective activation of only desired population of nerve fibres. Various neural interfaces have been already proposed for that purpose, including multipolar cuff electrodes, longitudinal intrafascicular electrodes (LIFE) and the Utah Slanted Electrode Array (USEA), all with different selectivity and invasiveness ratios. Recently a new electrode concept of a transversal intrafascicular multichannel electrode (TIME) has been proposed (Fig. 14). This electrode has been developed in frame of the European Project TIME in which the DEMAR team is participating (grant CP-FP-INFSO 224012 from the European Union). It is intended to be implanted transversally in the nerve and address several fascicles or subgroups of nerve fibres with one device. It has longitudinal shape and has several independent stimulation sites equally spread on both sides of the electrode.

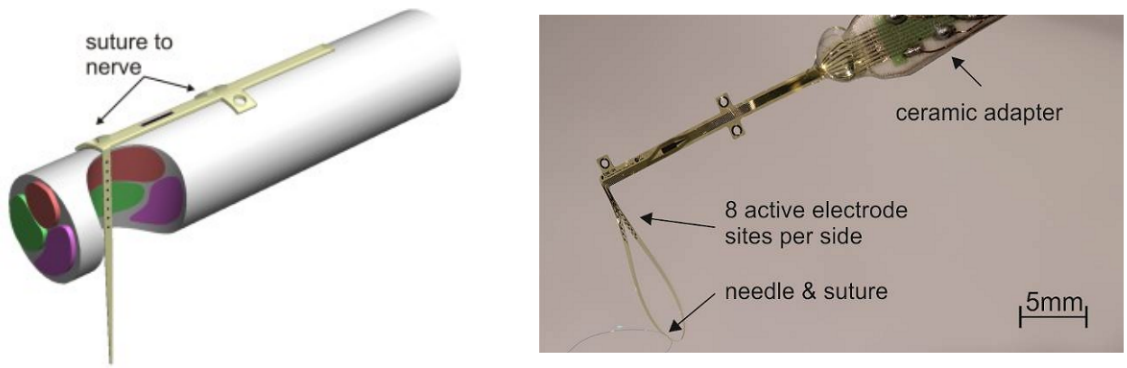


Figure 14. Left: The schematic representation of the TIME electrode implanted into the peripheral nerve. Right: detailed photograph of an TIME-3H electrode with pre-attached loop-thread (Source: Boretius et al. Proc. IEEE Biomedical Robotics and Biomechatronics Conference, Rome 2012)

It has been already shown that TIME allows to achieve high selectivity of stimulation when using monopolar configuration, i.e. when current is delivered through one of the sites of the TIME against small needle electrode placed in the proximity of the nerve. We have performed investigation in the sciatic nerve of rat to verify if the use of bipolar configuration, i.e. when current is delivered through one of the TIME sites against an other site of the same electrode, could allow to further enhance selectivity of stimulation. The results of our studies suggest that using bipolar configuration do allow to increase selectivity of stimulation. However, higher charge of the stimulation may be necessary to achieve similar level of muscle activation, as compared to the monopolar configuration [24].

When applied in the rat model the Transverse Intrafascicular Multi-channel Electrode (TIME) showed selective nerve fascicle recruitment. But results from the larger and poly-fasicular median nerves in pigs indicated that a single TIME could not reach the entire nerve and could only selectively recruit a subset of the nerve fasicles. The use of multiple TIME structures could offer a means to achieve highly selective fascicular stimulation while reaching a larger percentage of the fascicles in the nerve. Therefore we have investigated the use of pairs of TIMEs implanted in the median nerves of anesthesized pigs. TIME structures were implanted at different angles relative to each other or in parallel with one another. Electrical stimuli was passed through each contact of each TIME and the resulting electromyograms were recorded from seven muscles innervated by the median nerve. The ability to recruit these muscles was used to assess the stimulation selectivity of each contact using a selectivity index comparing the root-mean-square of the the evoked EMG of individul muscles. Results showed a signifcant increase in the selectivity index, when using two TIMEs compared to one. The optimal improvement was obeserved when TIMEs were placed in parallel to each other in such a way that they interfaced non-overlapping nerve regions [18].

6.3.4. Nerve model for ENG recording

Participants: Olivier Rossel, Guy Cathébras, Fabien Soulier, Serge Bernard.

In the context of selective electroneurogram recording, we showed last year the efficiency of a *small tripole* filtering (the distance between contacts is $375\,\mu\mathrm{m}$) thanks to simulated signals. This recording is locally sensitive and greatly increases the selectivity of the electrode. This year, we realized an experiment to verify the simulated results. Theoritical study of the *small tripole* sensitivity was realized for a single-fiber action potential (SFAP). We wanted to proceed in the same way for the experiment by trying to measure a SFAP.

However, actual biological SFAP would be hardly measurable by a *small tripole* in an in-vitro experiment. So, we decided to choose an approach based on an artificial axon. In this artificial model, every parameter is under control. The position of the fiber, the nodes of Ranvier, the position of the measuring electrode, as well as the involved currents are perfectly known. This allows us to implement and to estimate with accuracy the filtering realized by the *small tripole* with exactly the same configurations as in simulations. And by performing measures for several radial distances, we can verify the influence of this distance, in order to estimate the sensibility the *small tripole*. Moreover, we can increase the activity amplitude to be higher than in real fiber and then achieve a beneficial signal to noise ratio.

6.3.4.1. Method

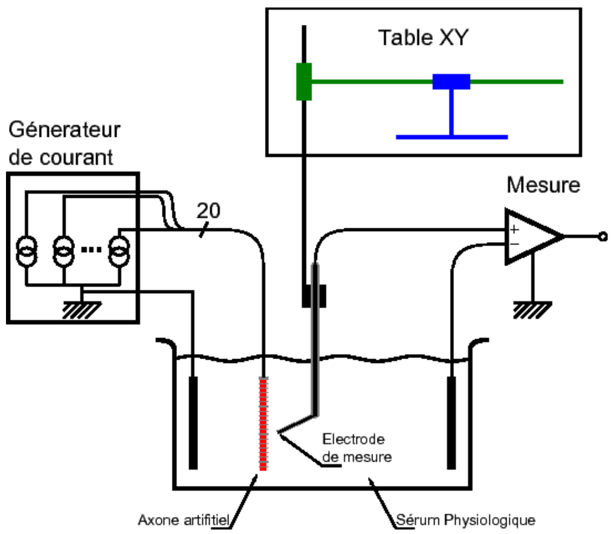


Figure 15. Principle of the experimental setup using the artificial axon.

The experimental setup consists in an artificial model emulating an axon and in a system measuring the action potential across the surrounding medium, as sketched in Fig. 15 . Biological tissues are modeled by a saline solution having a conductivity close to the human body. The space and time behavior of the current generated by a natural axon is reproduced on the artificial axon. The latter is emulated by an cochlear electrode. It exhibits 20 contacts, being an accurate image of the nodes of Ranvier on an axon of $8.7\,\mu\mathrm{m}$ diameter regarding the spatial periodicity.

In order to generate electric activity on several contacts that can be compared to the one of several nodes of Ranvier during the conduction of an action potential, we realized a custom multi-current generator with asynchronous outputs. The chosen amplitude was $35 \cdot 10^4$ times that of an human axon.

The measure of the SFAP in the space, is done using a punctual monopolar electrode. The position of the measuring electrode relative to the fiber is automatically set by a programmable micromanipulator. Then, the measure is repeated every $50\,\mu$m along $2\,mm$ on the radial axis.

6.3.4.2. Monopolar and tripolar sensitivity

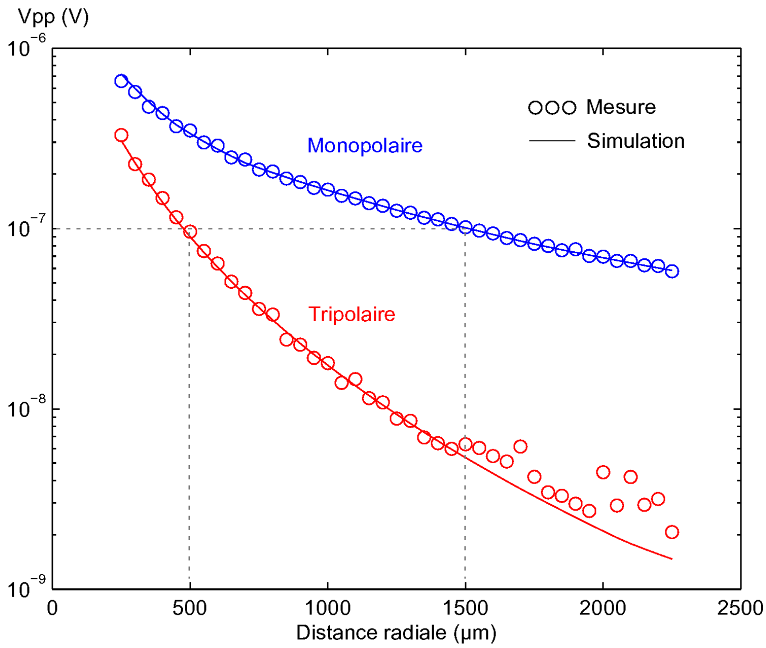


Figure 16. Comparison between monopolar and small-tripolar recording sensitivities. Measures are plotted with circles and simulation results are solid lines. The monopolar data are in blue and tripolar are in red.

To verify quantitatively the sensibility results obtained by simulation for monopolar and tripolar electrodes, we have first estimated the amplitude of a SFAP measured by a monopolar electrode. The measured SFAP amplitude directly gives the monopolar recording that varies according to the radial distance between the artificial axon and the electrode. Then, the *small-tripole* recording results from an off-line process combining three points of measure spaced out of $375\,\mu$m. The two kind of recordings are represented in Fig. 16 .

By comparing the recording sensitivities either estimated or measured, we can conclude that measures perfectly fit the theoretical results. We can also notice that the attenuation relative to the radial distance is more important for the *small tripole* than for the monopolar measure. This confirms that the measure performed by one *small tripole* electrode is exclusively sensitive to the closest fibers.

DRACULA Project-Team

6. New Results

6.1. Modelling of Erythroblastic Islands (red blood cell production)

Participants: Fabien Crauste [Contact person], Olivier Gandrillon, Vitaly Volpert.

In collaboration with N. Bessonov, S. Fischer and P. Kurbatova.

The production and regulation of red blood cells (erythropoiesis) occurs in the bone marrow where erythroid cells proliferate and differentiate within particular structures, called erythroblastic islands. A typical structure of these islands consists of a macrophage (white cell) surrounded by immature erythroid cells (progenitors), with more mature cells on the periphery of the island, ready to leave the bone marrow and enter the bloodstream.

We proposed a hybrid model [11], coupling a continuous model (ordinary differential equations) describing intracellular regulation through competition of two key proteins, to a discrete spatial model describing cell-cell interactions, with growth factor diffusion in the medium described by a continuous model (partial differential equations), to investigate the role of the central macrophage in normal erythropoiesis. Intracellular competition of the two proteins leads the erythroid cell either to proliferation, differentiation, or death by apoptosis. This approach allows considering spatial aspects of erythropoiesis, involved for instance in the occurrence of cellular interactions or the access to external factors, as well as dynamics of intracellular and extracellular scales of this complex cellular process, accounting for stochasticity in cell cycle durations and orientation of the mitotic spindle. The analysis of the model showed a strong effect of the central macrophage on the stability of an erythroblastic island, when assuming the macrophage releases prosurvival cytokines. Even though it is not clear whether or not erythroblastic island stability must be required, investigation of the model concludes that stability improves responsiveness of the model, hence stressing out the potential relevance of the central macrophage in normal erythropoiesis.

6.2. Modelling of the CD8 T cell Immune Response

Participants: Fabien Crauste [Contact person], Olivier Gandrillon, Emmanuelle Terry.

In collaboration with J. Marvel and C. Arpin.

The CD8 immune response is a specific immune response triggered by the organism when the innate response is unable to fight a pathogen. We proposed a new model of the CD8 T cell immune response based on the description of feedback controls exerted by the cytotoxic CD8 T cell population on the pathogen and the population itself [14]. This model, a system of ordinary and age-structured partial differential equations, allows describing a classical response, characterized by a cellular expansion following the pathogen-mediated activation, then a contraction phase and the generation of memory CD8 T cells. Moreover, we showed the global asymptotic stability of this system corresponding to the elimination of the virus. This situation is expected and describes for instance what is observed with the flu virus.

A simpler version of this model (based on nonlinear ordinary differential equations) has then been confronted to experimental data generated by Jacqueline Marvel's team in Lyon (immunology team), with 3 different pathogens. A parameter sweep has been performed and some parameters of the model, specific of cellular processes, have been shown to characterize CD8 immune responses against either a virus or a bacterium. This work is in progress and should be submitted soon.

6.3. Modelling of Platelet Thrombus Formation

Participants: Alen Tosenberger, Vitaly Volpert [Contact person].

In collaboration with F. Ataullakhanov, N. Bessonov, A. Butylin, G. Panasenko, M. Panteleev, E. Shnol, I. Sirakov and A. Tokarev.

An injury of a blood vessel requires quick repairing of the wound in order to prevent a loss of blood. This is done by the hemostatic system. The key point of its work is the formation of an aggregate from special blood elements, namely, platelets. The construction of a mathematical model of the formation of a thrombocyte aggregate with an adequate representation of its physical, chemical, and biological processes is an extremely complicated problem. A large size of platelets compared to that of molecules, strong inhomogeneity of their distribution across the blood flow, high shear velocities, the moving boundary of the aggregate, the interdependence of its growth and the blood flux hamper the construction of closed mathematical models convenient for biologists. We proposed a new PDE-based model of a thrombocyte aggregate formation [21], [22]. In this model, the movement of its boundary due to the adhesion and detachment of platelets is determined by the level set method. The model takes into account the distribution inhomogeneity of erythrocytes and platelets across the blood flow, the invertible adhesion of platelets, their activation, secretion, and aggregation. The calculation results are in accordance with the experimental data concerning the kinetics of the ADP-evoked growth of a thrombus in vivo for different flow velocities. The model constructed here can be easily extended to the case of other hemostatic mechanisms and can be integrated into different continuous blood flow models.

6.4. Reaction-Diffusion Model of Atherosclerosis Development

Participant: Vitaly Volpert [Contact person].

In collaboration with N. El Khatib, S. Genieys and B. Kazmierczak.

Atherosclerosis begins as an inflammation in blood vessel walls (intima). The inflammatory response of the organism leads to the recruitment of monocytes. Trapped in the intima, they differentiate into macrophages and foam cells leading to the production of inflammatory cytokines and further recruitment of white blood cells. This self-accelerating process, strongly influenced by low-density lipoproteins (cholesterol), results in a dramatic increase of the width of blood vessel walls, formation of an atherosclerotic plaque and, possibly, of its rupture. We suggested a 2D mathematical model of the initiation and development of atherosclerosis which takes into account the concentration of blood cells inside the intima and of pro- and anti-inflammatory cytokines [18]. The model represents a reaction-diffusion system in a strip with nonlinear boundary conditions which describe the recruitment of monocytes as a function of the concentration of inflammatory cytokines. We proved the existence of travelling waves described by this system and confirmed our previous results which suggest that atherosclerosis develops as a reaction-diffusion wave.

6.5. Hematopoietic model with feedback control

Participants: Mostafa Adimy [Contact person], Lila Sebaa.

In collaboration with O. Angulo and C. Marquet.

We investigate a mathematical model of blood cell production in the bone marrow (hematopoiesis). The model describes both the evolution of primitive hematopoietic stem cells and the maturation of these cells as they differentiate to form the three types of blood cells (red blood cells, white cells and platelets). The primitive hematopoietic stem cells and the first generations of each line (progenitors) are able to self-renew, and can be either in a proliferating or in a resting phase (G_0-phase). These properties are gradually lost while cells become more and more mature. The three types of progenitors and mature cells are coupled to each other via their common origin in primitive hematopoietic stem cells compartment. Peripheral control loops of primitive hematopoietic stem cells and progenitors as well as a local autoregulatory loop are considered in the model. The resulting system is composed by eleven age-structured partial differential equations. To analyze this model, we don't take into account cell age-dependence of coefficients, that prevents a usual reduction of the structured system to an unstructured delay differential system. We investigate some fundamental properties of the solutions of this system, such as boundedness and positivity. We study the existence of stationary solutions: trivial, axial and positive steady states. Then we give conditions for the local asymptotic stability of the trivial steady state and by using a Lyapunov function, we obtain a sufficient condition for its global asymptotic stability [7]. In some particular cases, we analyze the local asymptotic stability of the positive steady state by using the characteristic equation. Finally, by numerical simulations, we illustrate our results and we show that a change in the duration of cell cycle can cause oscillations. This can be related to observations of some periodical hematological disease such as chronic myelogenous leukemia, cyclical neutropenia, cyclical thrombocytopenia, etc.

DYLISS Team

6. New Results

6.1. Data integration

Participants: Jacques Nicolas [**contact**], Geoffroy Andrieux, Andres Aravena, Pierre Blavy, Jérémie Bourdon, Guillaume Collet, Damien Eveillard, Michel Le Borgne, Sylvain Prigent, Anne Siegel, Sven Thiele, Valentin Wucher.

- **Identification of key regulators by the integration of flux and regulatory information** [*P. Blavy, A. Siegel*] We introduced a new method to combine reaction-based "flux" information (consumption and prediction of molecules) and regulatory "causal" information (effect of the variation of a molecule on the variation of another molecule) in order to find potential key regulators of a set of molecules. It has been validated by recovering among the causal graph derived from the Transpath database the main regulators of 190 groups of genes which are known to share a transcription factor according to the TRED database. [22][Online publication]
- **Reconstruction of transcriptional networks** [*A. Aravena, A. Siegel*] Transcriptional regulatory network models can be reconstructed ab initio from DNA sequence data by locating the binding sites, defined by position specific score matrices, and identifying transcription factors by homology with known ones in other organisms. In general the resulting network contains spurious elements. We use differential expression experimental data, in the form of Mutual Information, as ASP logical constraints to be satisfied by any valid regulatory network subgraph. These rules are used to determine the minimal sets of motif and transcription factors which constitute a genetic regulatory network compatible with experimental data [20][Online publication].
- **Studying diversity in marine environment** [*D. Eveillard*] We proposed a statistical-based data analysis of environmental microarrays. It shows that similar physical parameters drive bacterial and archae communities that share common ammonia oxidizing capacities [12][Online publication]
- **Brown algae metabolic network reconstruction** [*S. Prigent, S. Thiele, A. Siegel*] In order to better understand the functioning of cellular metabolism in the model brown alga *E. siliculosus*, metabolic networks are under construction based on genomic information. Two approaches are conducted in parallel to complete the network, a stochastic one that proceeds by sampling the solution space and a combinatorial one that tries to minimize the number of added reactions [23].

6.2. Asymptotic dynamics

Participants: Anne Siegel [**contact**], Oumarou Abdou-Arbi, Geoffroy Andrieux, Pierre Blavy, Jérémie Bourdon, Damien Eveillard, Michel Le Borgne, Vincent Picard, Sven Thiele, Santiago Videla.

- **Probabilistic sources for sequences and systems biology** [*J. Bourdon*] The habilitation thesis surveys how methods based on average-case analysis of algorithms can be used to model the quantitative response of a biological system from a biomolecular to a physiological scale [28].
- **Learning the early-response of protein signaling networks.** [*S. Videla, S. Thiele, A. Siegel*] We demonstrated the usefulness of the Answer Set Programming approach (ASP) to learn Boolean models from high-throughput phospho-proteomics data. Exact constraint solving showed a quantum leap over heuristic (state-of-the-art) methods in terms of efficiency and scalability, and guarantees global optimality of solutions as well as provides a complete set of solutions [19][Online publication]
- **Numerical model of signaling pathways** [*G. Andrieux, M. Le Borgne*] We have proposed an integrative numerical (ODE) model for the dynamic regulation of $TGF\beta$ Signaling by $TIF1\gamma$. The model successfully unifies the seemingly opposite roles of $TIF1\gamma$, and reveals how changing $TIF1\gamma/Smad4$ ratios affect the cellular response to stimulation by $TGF\beta$, accounting for a highly graded determination of cell fate. [10].

- **Identification of regulatory networks in ecology** [*D. Eveillard*] A clustering data-based approach emphasizes regulatory networks at the bacterial population scale. It allowed the identification of antagonistic interactions between heterotrophic bacteria as a potential regulator of community structure of hypersaline microbial mats. [15][Online publication]

6.3. Sequence annotation

Participants: François Coste [**contact**], Catherine Belleannée, Gaëlle Garet, Clovis Galiez, Laurent Miclet, Jacques Nicolas.

- **Expressive pattern matching** [*C. Belleannée, J. Nicolas*] We have presented for the first time Logol, a new application designed to achieve pattern matching in possibly large sequences with realistic biological motifs. Logol consists in both a language for describing patterns, and the associated parser for effectively scanning sequences (RNA, DNA or protein) with such motifs. The language, based on an high level grammatical formalism, allows to express flexible patterns (with misparings - improper alignment of DNA strands - and indels) composed of both sequential and structural elements (such as repeats or pseudoknots)[21][Online publication]. Logol has been applied to the detection of -1 frameshifts, a structure including pseudoknots, on a reference benchmark (Recode2) [26][Online publication].
- **Analysis of sequence repeats** [*J. Nicolas*] We have participated to a book that introduces up-to-date methods for the identification and study of transposable elements in genomes. J. Nicolas contributed with a chapter that provides an overview of the formal underpinnings of the search for these highly repeated elements in genomic sequences and describes a selection of practical tools for their analysis. It concludes with the interest of syntactic analysis in this domain [24][Online publication].
- **Grammatical models for local patterns** [*G. Garet, J. Nicolas, F. Coste*] We studied the annotation of new proteins with respect to banks of already annotated protein sequences. For this task, we are developping grammatical inference methods. We introduced new classes of substitutable languages and new generalization criterion based on local substitutability concept and illustrated the great potential of the approach on a benchmark considering a real non trivial protein family. [16][Online publication]
- **Local maximality** [*L. Miclet*] Starting from locally maximal subwords and locally minimal superwords common to a finite set of words, we have defined the corresponding sets of alignments. We gave a partial order relation between such sets of alignments, as well as two operations between them and showed it has a lattice structure that can be used for inducing a generalization of the set of words [18][17].
- **Searching for Smallest Grammars on Large Sequences and Application to DNA** [*F. Coste*] We are motivated by the inference of the structure of genomic sequences, that we address as an instance of the smallest grammar problem. Previously, we reduce it to two independent optimization problems: choosing which words will be constituents of the final grammar and finding a minimal parsing with these constituents. This year we made these ideas applicable on large sequences. First, we improved the complexity of existing algorithms by using the concept of maximal repeats for constituents. Then, we improved the size of the grammars by cautiously adding a minimal parsing optimization step. Together, these approaches enabled us to propose new practical algorithms that return smaller grammars (up to 10%) in approximately the same amount of time than their competitors on a classical set of genomic sequences and on whole genomes. [14] [Online publication].
- **CyanoLyase: a database of phycobilin lyase sequences, motifs and functions** [*F. Coste*] In collaboration with our partners of the ANR project Pelican, we have set up CyanoLyase (http://cyanolyase.genouest.org/), a manually curated sequence and signature database of phycobilin lyases and related proteins. Protomata-Learner has been used to establish the signature of the 32 known subfamilies that are used to rapidly retrieve and annotate lyases from any new genome [13] [Online publication]

FLUMINANCE Project-Team

6. New Results

6.1. Fluid motion estimation

6.1.1. Stochastic uncertainty models for motion estimation

Participants: Thomas Corpetti, Etienne Mémin.

In this study we have proposed a stochastic formulation of the brightness consistency used principally in motion estimation problems. In this formalization the image luminance is modeled as a continuous function transported by a flow known only up to some uncertainties. Stochastic calculus then enables to built conservation principles which take into account the motion uncertainties. These uncertainties defined either from isotropic or anisotropic models can be estimated jointly to the motion estimates. Such a formulation besides providing estimates of the velocity field and of its associated uncertainties allows us to define a natural linear scale space multiresolution framework. The corresponding estimator, implemented within a local least squares approach, has shown to improve significantly the results of the corresponding deterministic estimator (Lucas and Kanade estimator). This fast local motion estimator has been shown to provide results that are of the same order of accuracy than state-of-the-art dense fluid flow motion estimator for particle images. The uncertainties estimated provide a useful piece of information in the context of data assimilation. This ability has been exploited to define multiscale data assimilation filtering schemes. These works have been recently published in *IEEE trans. on Image Processing* and in Numerical Mathematics: Theory, Methods and Applications [16], [18]. We intend to pursue this formalization to define dense motion estimators that allow us handling, in the same way, luminance conservation under motion uncertainty principles. An efficient GP-GPU implementation of the local estimator is also targeted.

6.1.2. 3D flows reconstruction from image data

Participants: Ioana Barbu, Cédric Herzet, Etienne Mémin.

Our work focuses on the design of new tools for the problem of 3D reconstruction of a turbulent flow motion. This task includes both the study of physically-sound models on the observations and the fluid motion, and the design of low-complexity and accurate estimation algorithms. On the one hand, state-of-the-art methodologies such as "sparse representations" will be investigated for the characterization of the observation and fluid motion models. Sparse representations are well-suited to the representation of signals with very few coefficients and offer therefore advantages in terms of computational and storage complexity. On the other hand, the estimation problem will be placed into a probabilistic Bayesian framework. This will allow the use of state-of-the-art inference tools to effectively exploit the strong time-dependence of the fluid motion. In particular, we will investigate the use of "ensemble Kalman" filter to devise low-complexity sequential estimation algorithms.

At the beginning of Ioana Barbu's PhD, we concentrated our efforts on the problem of reconstructing the particle positions from several two-dimensional images. Our approach is based on the exploitation of a particular family of sparse representation algorithms, namely the so-called "pursuit algorithms". Indeed, the pursuit procedures generally allow a good trade-off between performance and complexity. Hence, we have performed a thorough study comparing the reconstruction performance and the complexity of different state-of-the-art algorithms to that achieved with pursuit algorithms. This work has led to the publication of two conference papers in experimental fluid mechanics.

This year, our work has focused on: i) the estimation of the 3D velocity field of the fluid flow from the reconstructed volumes of particles; ii) the design of new methodologies allowing to jointly estimate the volume of particles and the velocity field from the received image data. More particularly, we have implemented a motion estimator generalizing the local Lucas-Kanade's procedure to a 3D problem. A potential strength of the proposed approach is the possibility to consider a fully parallel (and therefore very fast) implementation. On the other hand, we have started investigating the problem of jointly estimating the volumes of particles and the velocity field. Our approach is based on the combination of sparse representation algorithms and "Lucas-Kanade"-like motion estimation methods. We are about testing the proposed approach on experimental data in order to assess its performance in practical scenarios of fluid mechanics. We also intend to collaborate with the group of Fulvio Scarano at TU Delft to assess and compare our method on experimental 3D data.

6.1.3. Motion estimation techniques for turbulent fluid flows

Participants: Patrick Héas, Dominique Heitz, Cédric Herzet, Etienne Mémin.

Based on physical laws describing the multi-scale structure of turbulent flows, this study concerns the proposition of smoothing functional for the estimation of homogeneous turbulent flow velocity fields from an image sequence. This smoothing is achieved by imposing some scale invariance property between histograms of motion increments computed at different scales. By reformulating this problem from a Bayesian perspective, an algorithm is proposed to jointly estimate motion, regularization hyper-parameters, and to select the most likely physical prior among a set of models. Hyper-parameter and model inference is conducted by likelihood maximization, obtained by marginalizing out non-Gaussian motion variables. The Bayesian estimator is assessed on several image sequences depicting synthetic and real turbulent fluid flows. Results obtained with the proposed approach in the context of fully developped turbulence improve significantly the results of state of the art fluid flow dedicated motion estimators. This series of works, which have been done in close collaboration with P. Minnini (University of Buenos Aires), have been published in several journals [21], [22], [23].

6.1.4. Wavelet basis for multi-scale motion estimation

Participants: Pierre Dérian, Patrick Héas, Cédric Herzet, Souleymane Kadri Harouna, Etienne Mémin.

This work describes the implementation of a simple wavelet-based optical-flow motion estimator dedicated to the recovery of fluid motion. The wavelet representation of the unknown velocity

field is considered. This scale-space representation, associated to a simple gradient-based optimization algorithm, sets up a natural multiscale/multigrid optimization framework for the optical flow estimation that can be combined to more traditional incremental multiresolution approaches. Moreover, a very simple closure mechanism, approximating locally the solution by high-order polynomials, is provided by truncating the wavelet basis at intermediate scales. This offers a very interesting alternative to traditional Particle Image Velocimetry techniques. As another alternative to this medium-scale estimator, we explored strategies to define estimation at finer scales. These strategies rely on the encoding of high-order smoothing functional on appropriate wavelet basis. Divergence-free bi-othogonal wavelet bases enable to further nicely enforce volume preserving motion field. Numerical results on several examples have demonstrated the relevance of the method for divergence free-2D flows. These studies have been published in the journal of Numerical Mathematics: Theory, Methods and Applications [19] and in the journal of Computer Vision [24]. The extension to 3D flows would be an interesting perspective.

6.1.5. Wavelet-based divergence-free fBm prior: application to turbulent flow estimation

Participant: Patrick Héas.

This work is concerned with the estimation of turbulent flows from the observation of an image sequence. From a Bayesian perspective, we propose to study divergence-free isotropic fractional Brownian motion (fBm) as a prior model for instantaneous turbulent velocity fields. These priors are self-similar stochastic processes, which characterize accurately second-order statistics of velocity fields in incompressible isotropic turbulence. Although, these models belong to a well-identified family of rotation invariant regularizers, there is a lack of effective algorithms in the literature to deal in practice with their fractional nature. To respond to this problem, we propose to decompose fBms on well-chosen wavelet bases. As a first alternative, we propose to design wavelets as whitening filters for divergence-free isotropic fBms, which are correlated both in space and scale. The second alternative is to use a divergence-free wavelet basis, which will take implicitly into account the divergence-free constraint arising form the physics.

6.1.6. Sparse-representation algorithms

Participant: Cédric Herzet.

The paradigm of sparse representations is a rather new concept which appears to be central in many field of signal processing. In particular, in the field of fluid motion estimation, sparse representation appears to be potentially useful at several levels: i) it provides a relevant model for the characterization of the velocity field in some scenarios; ii) it turns out to be central in the recovering of volumes of particles in the 3D Tomo-PIV problem. In these contexts, the dimensionality of the problem can be very large and the use of sparse-representation algorithms allowing for a good trade-off between complexity and effectiveness is needed.

This year, we have therefore pursued our study of efficient sparse decomposition algorithms. In particular, we have extended our work addressing the problem of finding good sparse representations into a probabilistic framework. First, we have proposed a new family of pursuit algorithms able to take into account any type of dependence (e.g. spatial or temporal) between the atoms of the sparse decomposition. This work has led to the publication of a paper in the proceedings of the international conference IEEE ICASSP 2012.

Exploiting further this probabilistic framework, we have then considered the design of structured soft pursuit algorithms. In particular, instead of making hard decisions on the support of the sparse representation and the amplitude of the non-zero coefficients, our soft procedures iteratively update probability on the latter values. The proposed algorithms are designed within the framework of the mean-field approximations and resort to the so-called variational Bayes EM algorithm to implement an efficient minimization of a Kullback-Leibler criterion. On the other hand, the proposed methodologies can handle "structured" sparse representations, that is, sparse decompositions where some dependence exists between the non-zero elements of the support. The prior model on the support of the sparse decomposition is based on a Boltzmann machine which encompasses as particular cases many type of dependence (Markov chain, Ising model, tree-like structure, etc). This work has been published in the journal IEEE Trans. on Signal Processing in 2012.

6.2. Tracking and data assimilation

6.2.1. Stochastic filtering for fluid motion tracking

Participants: Sébastien Béyou, Anne Cuzol, Etienne Mémin.

We investigated the study of a recursive Bayesian filter for tracking velocity fields of fluid flows. The filter combines an Ito diffusion process associated to 2D vorticity-velocity formulation of Navier-Stokes equation and discrete image error reconstruction measurements. In contrast to usual filters designed for visual tracking problems, our filter combines a continuous law for the description of the vorticity evolution with discrete image measurements. We resort to a Monte-Carlo approximation based on particle filtering. The designed tracker provides a robust and consistent estimation of instantaneous motion fields along the whole image sequence.

When the likelihood of the measurements can be modeled as a Gaussian law, we have also investigated the use of the so-called ensemble Kalman filtering for fluid tracking problems. This kind of filters introduced for the analysis of geophysical fluids is based on the Kalman filter update equation. Nevertheless, unlike traditional Kalman filtering setting, the covariances of the estimation errors, required to compute the so-called Kalman gain, relies on an ensemble of forecasts. Such a process gives rise to a Monte-Carlo approximation for a family of non-linear stochastic filters enabling to handle state spaces of large dimension. We have recently proposed an extension of this technique that combines sequential importance sampling and the propagation law of an ensemble Kalman filter. This technique leads to an ensemble Kalman filter with an improved efficiency. We have in particular investigated the introduction of a nonlinear direct image measurement operator within this ensemble Kalman scheme. This modification of the filter provides very good results on 2D numerical and experimental flows even in the presence of strong noises. We are currently assessing its application to oceanic satellite images for the recovering of ocean streams. We are also studying the impact on the stochastic dynamics of turbulent noise defined as auto-similar Gaussian random fields and the introduction within an incremental ensemble analysis scheme of multiscale motion measurements. This work has been recently accepted for publication in the Tellus A journal [17].

6.2.2. Reduced-order models for flows representation from image data

Participants: Cédric Herzet, Etienne Mémin, Véronique Souchaud.

One of the possibilities to neglect the influence of some degrees of freedom over the main characteristics of a flow consists in representing it as a sum of K orthonormal spatial basis functions weighted with temporal coefficients. To determine the basis function of this expansion, one of the usual approaches relies on the Karhunen-Loeve decomposition (refered to as proper orthogonal decomposition – POD – in the fluid mechanics domain). In practice, the spatial basis functions, also called modes, are the eigenvectors of an empirical auto-correlation matrix which is built from “snapshots" of the considered physical process.

In this axis of work we focus on the case where one does not have a direct access to snapshots of the considered physical process. Instead, the POD has to be built from the partial and noisy observation of the physical phenomenon of interest. Instances of such scenarios include situations where real instantaneous vector-field snapshots are estimated from a sequence of *images*. We have been working on several approaches dealing with such a new paradigm. A first approach consists in extending standard penalized motion-estimation algorithms to the case where the sought velocity field is constrained to span a low-dimensional subspace. In particular, we have considered scenarios where the standard optical flow constraint (OFC) is no longer statisfied and one has therefore to resort to a Discrete Finite Difference (DFD) model. The non-linearity of the latter leads to several practical issues that we have addressed this year. We are currently assessing the performance of the proposed method on experimental data in order to validate its relevance in practical scenarios. In a second approach we have studied two variational data assimilation techniques for the estimation of low order dynamical models for fluid flows. Both methods are built from optimal control recipes and rely on POD representation associated to Galerkin projection of the Navier Stokes equations. The proposed techniques differ in the control variables they involve. The first one introduces a weak dynamical model defined only up to an additional uncertainty time dependent function whereas the second one, handles a strong dynamical constraint in which the coefficients of the dynamical system constitute the control variables. Both choices correspond to different approximations of the relation between the reduced basis on which is expressed the motion feld and the basis components that have been neglected in the reduced order model construction. The techniques have been assessed on numerical data and for real experimental conditions with noisy Image Velocimetry data. This work has been published in the Journal of Computational Physics [15]. In collaboration with the University of Buenos Aires, we have also explored, a method that combines Proper Orthogonal Decomposition with a spectral technique to analyze and extract reduced order models of flows from time resolved data of velocity fields. This methodology, relying on the eigenfunctions of the Koopman operator, is specifically adapted to flows with quasi periodic orbits in the phase space. The technique is particularly suited to cases requiring a discretization with a high spatial and temporal resolution. The proposed analysis enables to decompose the flow dynamics into modes that oscillate at a single frequency. For each modes an energy content and a spatial structure can be put in correspondence. This approach has been assessed for a wake flow behind a cylinder at Reynolds number 3900 and has been recently accepted under minor revisions condition to the journal of Theoretical and Computational Fluid Dynamics.

6.2.3. Optimal control techniques for the coupling of large scale dynamical systems and image data

Participants: Dominique Heitz, Etienne Mémin, Cordelia Robinson, Yin Yang.

This work aims at investigating the use of optimal control techniques for the coupling of Large Eddies Simulation (LES) techniques and 2D image data. The objective is to reconstruct a 3D flow from a set of simultaneous time resolved 2D image sequences visualizing the flow on a set of 2D plans enlightened with laser sheets. This approach will be experimented on shear layer flows and on wake flows generated on the wind tunnel of Irstea Rennes. Within this study we wish also to explore techniques to enrich large-scale dynamical models by the introduction of uncertainty terms or through the definition of subgrid models from the image data. This research theme is related to the issue of turbulence characterization from image sequences. Instead of predefined turbulence models, we aim here at tuning from the data the value of coefficients involved in traditional LES subgrid models or in longer-term goal to learn empirical subgrid models directly from image data. An accurate modeling of this term is essential for Large Eddies Simulation as it models all the non resolved motion scales and their interactions with the large scales.

We have pursued the first investigations on a 4DVar assimilation technique, integrating PIV data and Direct Numerical Simulation (DNS), to reconstruct two-dimensional turbulent flows. The problem we are dealing with consists in recovering a flow obeying Navier-Stokes equations, given some noisy and possibly incomplete PIV measurements of the flow. By modifying the initial and inflow conditions of the system, the proposed method reconstructs the flow on the basis of a DNS model and noisy measurements. The technique has been evaluated in the wake of a circular cylinder. It denoises the measurements and increases the spatiotemporal resolution of PIV time series. These results have been conditionally accepted for publication in Journal of Computational Physics. Along the same line of studies we have started to investigate the 3D case. The goal consists here to reconstruct a 3D flow from a set of simultaneous time resolved 2D images of planar sections of the 3D volume. This work is mainly conducted within the PhD of Cordelia Robinson. The development of the variational assimilation code has been initiated within a collaboration with A. Gronskis, S. Laizé (lecturer, Imperial College, UK) and Eric Lamballais (institut P' Poitiers).

6.2.4. Free surface flows reconstruction and tracking

Participants: Benoît Combes, Dominique Heitz, Etienne Mémin, Cordelia Robinson, Yin Yang.

Characterizing a free-surface flow (space and time-dependent velocity and geometry) given observations/measures at successive times is an ubiquitous problem in fluid mechanic and in hydrology. Observations can consist of e.g. measurements of velocity, or like in this work of measurements of the geometry of the free-surface. Indeed, recently developed depth/range sensors allow to capture directly a rough 3D geometry of surfaces with high space and time resolution. We have investigated the performance of the Kinect and have shown that it is likely to capture temporal sequences of depth observations of wave-like surfaces with wavelengths and amplitudes sufficiently small to characterize medium/large scale flows. Several data assimilation methods have been experimented and compared to estimate both time dependent geometry and displacement field associated to a free-surface flow from a temporal sequence of Kinect data. This study have been conducted on synthetic and real-world data. It has been presented to a data assimilation conference [35]. Finally, we explored the application of such techniques to hydrological applications. These results are currently considered for submission to Journal of Hydrology.

6.2.5. Stochastic filtering technique for the tracking of closed curves

Participants: Christophe Avenel, Etienne Mémin.

We have proposed a filtering methodology for the visual tracking of closed curves. Opposite to works of the literature related to this issue, we consider here a curve dynamical model based on a continuous time evolution law with different noise models. This led us to define three different stochastic differential equations that capture the uncertainty relative to curve motions. This new approach provides a natural understanding of classical level-set dynamics in terms of such uncertainties. These evolution laws have been combined with various color and motion measurements to define probabilistic state-space models whose associated Bayesian filters can be handled with particle filters. This ongoing work will be continued within extensive curve tracking experiments and extended to the tracking of other very high dimensional entities such as vector fields and surfaces. This work, which corresponds to the PhD thesis of Christoph Avenel has been presented in several conferences and has been submitted to two different journals. Let us note that it has also led to a fruitful collaboration with MeteoFrance [30]

6.2.6. Sequential smoothing for fluid motion

Participants: Anne Cuzol, Etienne Mémin.

In parallel to the construction of stochastic filtering techniques for fluid motions, we have proposed a new sequential smoothing method within a Monte-Carlo framework. This smoothing aims at reducing the temporal discontinuities induced by the sequential assimilation of discrete time data into continuous time dynamical models. The time step between observations can indeed be long in environmental applications for instance, and much longer than the time step used to discretize the model equations. While the filtering aims at estimating the state of the system at observations times in an optimal way, the objective of the smoothing is to improve the estimation of the hidden state between observation times. The method is based on a Monte-Carlo approximation of the filtering and smoothing distributions, and relies on a simulation technique of conditional diffusions. The proposed smoother can be applied to general non linear and multidimensional models. It has been applied to a turbulent flow in a high-dimensional context, in order to smooth the filtering results obtained from a particle filter with a proposal density built from an Ensemble Kalman procedure. This conditional simulation framework can also be used for filtering problem with low measurement noise. This has been explored through a collaboration with Jean-Louis Marchand (LORIA) in the context of vorticity tracking from image data.

6.2.7. Stochastic fluid flow dynamics under uncertainty

Participant: Etienne Mémin.

In this research axis we aim at devising stochastic Eulerian expressions for the description of fluid flow evolution laws incorporating uncertainty on the particles location. Such an uncertainty modeled through the introduction of a random term, allows taking into account approximations or truncation effects performed within the dynamics analytical constitution steps. This includes for instance the modeling of unresolved scales interaction in large eddies simulation (LES) or in Reynolds average numerical simulation (RANS), but also uncertainties attached to non-uniform grid discretization. This model is mainly based on a stochastic version of the Reynolds transport theorem. Within this framework various simple expressions of the mean drift component can be exhibited for different models of the random field carrying the uncertainties we have on the flow. We aim at using such a formalization within image-based data assimilation framework and to derive appropriate stochastic versions of geophysical flow dynamical modeling.

6.2.8. Variational assimilation of images for large scale fluid flow dynamics with uncertainty

Participants: Souleymane Kadri Harouna, Etienne Mémin.

In this work we explore the assimilation of a large scale representation of the flow dynamics with image data provided at a finer resolution. The velocity field at large scales is described as a regular smooth components whereas the complement component is a highly oscillating random velocity field defined on the image grid but living at all the scales. Following this route we have started to assess the performance of a variational assimilation technique with direct image data observation. Preliminary encouraging results obtained for a wavelet-based 2D Navier Stokes implementation and images of a passive scalar transported by the flow have been obtained. Large-scale simulation under uncertainty for the 3D viscous Taylor-Green vortex flow have been carried out and show promising results of the approach.

6.3. Analysis and modeling of turbulent flows

6.3.1. Hot-wire anemometry at low velocities

Participant: Dominique Heitz.

A new dynamical calibration technique has been developed for hot-wire probes. The technique permits, in a short time range, the combined calibration of velocity, temperature and direction calibration of single and multiple hot-wire probes. The calibration and measurements uncertainties were modeled, simulated and controlled, in order to reduce their estimated values. Based on a market study the french patent application has been extended this year to a Patent Cooperation Treaty (PCT) application.

6.3.2. Numerical and experimental image and flow database

Participant: Dominique Heitz.

The goal was to design a database for the evaluation of the different techniques developed in the Fluminance group. The main challenge was to enlarge a database mainly based on two-dimensional flows, with three-dimensional turbulent flows. New synthetic image sequences based on homogeneous isotropic turbulence and on circular cylinder wake have been provided. These images have been completed with real image sequences based on wake and mixing layers flows. This new database provides different realistic conditions to analyse the performance of the methods: time steps between images, level of noise, Reynolds number, large-scale images.

6.4. Visual servoing approach for fluid flow control

6.4.1. Fully exploitation of the controlled degrees of freedom of the 2D plane Poiseuille flow

Participants: Christophe Collewet, Xuan Quy Dao.

This works concerns the Phd of Xuan-Quy Dao and can be seen as an extension of the works carried out by Romeo Tatsambon during its post-doc position. Indeed, during this post-doc we proved that our vision-based approach overcomes the traditional approaches. Nevertheless, to compare our method with the literature, we used a traditional control law, the LQR control law. However, we can fully exploit the capabilities of visual servoing techniques by designing a more efficient control law than the LQR one. This has been done this year. We have validated our approach to the problem of minimizing the drag of the 2D plane Poiseuille flow. An important issue was also to ensure that, during the process of drag reduction, the kinetic energy density will not grow. This is of great importance since it is well known that the controlled flow may become turbulent when this kinetic energy density is growing. To cope with this problem we have proposed to design a control law based on partitioned visual servo. Indeed, following this way, we are able to simultaneously minimize the drag AND the kinetic energy density in contrast to the existing approaches. This work has been accepted to the "American control conference (ACC'12)", to the "Conférence internationale francophone d'automatique (CIFA'12)" and to the "6th AIAA Flow Control Conference". We have also explored an approach based on eigenstructure placement to ensure a strict decrease of the kinetic energy density. Another approach has also been explored, it tends to decouple all the controlled degrees of freedom of the system so that an exponential decoupled decrease of each component of the state vector is obtained. The great advantage of this approach is that all quantities depending on the state vector (like the drag) are also exponential decreasing functions.

6.4.2. Control behind a backward-facing step

Participant: Christophe Collewet.

Instead of setting up an experimental closed loop control problem for the plane Poiseuille flow with temporal perturbations, which is theoretically based on an unrealistic infinite channel, we explore in this axis of work a closed-loop control of a flow behind a backward step. The control is expressed through the visual servoing formalism and fast velocity measurements in the recirculation zone. This work is performed in the context of the PhD thesis of Nicolas Gautier from PMMH-ESCPI. This thesis is co-supervised with Jean-Luc Aider (CNRS/PMMH-ESPCI).

GALEN Team

6. New Results

6.1. Machine Learning & Optimization

Participants: Andreas Argyriou, Matthew Blaschko, Pawan Kumar.

- **Sparse Prediction & Convex Optimization Decomposition** [*Andreas Argyriou*]
  In [36], we have introduced a new regularization penalty for sparse prediction, the k-support norm. This norm corresponds to the tightest convex relaxation of sparsity combined with an ℓ_2 penalty. We have shown that this new norm provides a tighter relaxation than the elastic net, and is thus a good replacement for the Lasso or the elastic net in sparse prediction problems. In [41], motivated by learning problems we proposed a novel optimization algorithm for minimizing a convex objective which decomposes into three parts: a smooth part, a simple non-smooth Lipschitz part, and a simple non-smooth non-Lipschitz part.
- **Learning Optimization for NP-complete Inference** [*Matthew Blaschko*]
  In [14] an optimization strategy for learning to optimize boolean satisfiability (SAT) solvers is given. Applications to real-world SAT problems show improved computational performance as a result of the learning algorithm.
- **Max-Margin Min-Entropy Models & Dissimilarity Coefficient based Learning** [*Pawan Kumar*]
  In [22] we proposed the family of max-margin min-entropy (M3E) models, which predicts a structured output for a given input by minimizing the Renyi entropy. The parameters of M3E are learned by minimizing an upper bound on a user-defined loss. We demonstrated the efficacy of M3E on two problems using publicly available datasets: motif finding and image classification. In [19] we proposed a novel structured prediction framework for weakly supervised datasets. The framework minimizes a dissimilarity coefficient between the predictor and a conditional distribution over the missing information. We demonstrated the efficacy of our approach on two problems using publicly available datasets: object detection and action detection.

6.2. Computational Vision & Perception

Participants: Matthew Blaschko, Iasonas Kokkinos, Pawan Kumar, Nikos Paragios.

- **Structured Output Ranking & Detailed Understanding of Objects in Computer Vision** [*Matthew Blaschko*]
  In [23] we proposed a novel method for efficiently optimizing an objective that ranks structured outputs by their loss. Based on the observation that structured output spaces [9] in computer vision problems can be well-modeled by a small number of loss values, our algorithm is able to optimize a quadratic number of pairwise constraints in linear time. In [38] we detail the research activities of a summer workshop hosted by Johns Hopkins University on learning a detailed understanding of objects and scenes in natural images. We worked on automatic verification of annotations provided through Amazon Mechanical Turk [35], texture categorization, and dependence modeling for bottom up proposals.
- **Efficient inference and learning for structured probabilistic models of deformable objects** [*Iasonas Kokkinos*, Haithem Boussaid & Stavros Tsogkas]

We have developed novel features to describe surface points intrinsically through the Intrinsic Shape Context (ISC) descriptor published in [17]. This method has delivered state-of-the-art results in surface point matching and we will explore its use for surface correspondence. The implementation of these descriptors is publicly available. In [32] we proposed a learning-based approach to symmetry detection by fusing multiple cues related to image intensity, color and texture, which delivered state-of-the-art results. We intend to extend this approach to 3D image analysis, and in particular for medical images. The implementation of these detectors is publicly available. In [27] we introduce a grouping-based method to learn and detect action classes in spatio-temporal data. Our method can both classify actions and indicate the spatio-temporal structures which provide support for the decision. The implementation of our front-end is publicly available. In [40] we have extended our work on efficient algorithms for object detection to accommodate fast methods for computing the part scores in a principled optimization framework, while he have thoroughly presented it in [40] and made the implementation publicly available.

- **Multi-view Image Segmentation & Parsing** [*Nikos Paragios*]
  In [28] a method for image matching was proposed that exploits hierarchical image representations through higher order graphs. The matching was achieved through a graph-based theoretical framework where the similarity and spatial consistency of the image semantic objects is encoded in a graph of commute times that is also endowed with singleton terms through shape descriptors. Many-to-many matching of regions are specially challenging due to the instability of the segmentation under slight image changes, and we explicitly handle it through high order potentials. These ideas were further explored in the context of co-segmentation [29] where a method to determine a consistent partition of multiple images was introduced through a multi-scale multiple-image generative model based on region matching that exploits inter-image information and establishes correspondences between the common objects that appear in the scene. Last, but not least in [24] a method that combines bottom up (visual information, visual descriptors, elements detection) information and top-town models (hierarchical shape grammars) was considered towards automatic facade parsing though reinforcement learning while in [30] a method for 3D image parsing was proposed based on a hierarchical grammar that was performing explicit 3D modeling of the scene through a combination of multi-image segmentation and a depth reconstruction process. The problem optimal combination of these two concurrent terms was addressed trough a pareto-driven criterion while the optimization was addressed through an evolutionary computation algorithm.

6.3. Biomedical Image Analysis

Participant: Nikos Paragios.

- **Image Reconstruction** [*Nikos Paragios* & Hellene Langet]
  In [21] a novel iterative reconstruction algorithm based on compressed sensing was proposed for Digital Subtraction Rotational Angiography (DSRA) that exploits both spatial and temporal sparsity through a proximal implementation that accommodates multiple$L-1$-penalties. These ideas was further explored in [20] where we introduced a three-dimensional reconstruction of tomographic acquisitions in C-arm-based rotational angiography was proposed that was able to deal with the temporal variations due to intra-arterial injections through a compressed-sensing approach leading to significant motion artifacts reduction in spite of the cone-beam geometry, the short-scan acquisition, and the truncated and subsampled data.
- **Image Segmentation** [*Nikos Paragios*, Pierre-Yves Baudin, Xiang Bo & Sarah Parisot]
  In [11] the problem of human skeletal muscle segmentation was considered through a graph-based approach (random walker). An automatic seed placement framework was introduced through a graph-theoretic formulation. Towards accounting for anatomical constraints, the Random Walker algorithm was endowed with a liner sub-space statistical prior towards improving segmentation robustness on missing and incomplete data [12]. The same formulation was extended to cope with non-linear priors through a Gaussian-like local prior model penalizing the deviations of the

coefficients of the random walker diffusion matrix from the ones learned from the training data [13]. In [25] a novel graph-based prior was considered towards modeling the distribution of low-glioma brain tumors and spatially characterizing them through a sparse hierarchical graph. Such a prior model was integrated to an image-driven voxel-like segmentation framework where image separation was achieved through a machine learning method towards automatic detection, characterization and segmentation of brain tumors. Furthermore, towards encoding pose invariance in the context of knowledge-based segmentation in [33] where a higher order graph-based implicit pose invariant formulation was introduced for cardiac segmentation. The formulation was endowed with higher order cliques allowing (i) the estimation of boundary and regional image support and (ii) the implicit modeling of local deformations with respect to a prior statistical model while being invariant to linear transformations.

- **Image Registration** [*Nikos Paragios*, Nicolas Honnorat & Sarah Parisot]
  In [15] the problem of organ-driven registration was addressed through simultaneous combined fusion of multi-modal images in the context of guide-wire segmentation through fluoroscopic and contrast enhanced images. To this end, a graphical model was considered that was segmenting and registering the guide-wire in the two modalities while establishing correspondences between the associated curves as well. Similar philosophy was used in the [26] where a method for one shot deformable brain registration and tumor segmentation was proposed between a healthy anatomical atlas and a diseased patient. Both tasks were addressed through a discrete formulation (pair-wise MRF using grid-like deformation models and machine learning discriminative frameworks for the separation of healthy versus diseased tissues) while interconnections between the two graphs were used to alleviate the registration requirement on tumor areas. The problem of symmetric registration was studied in [31] through a common grid deforming in both directions according to a symmetric manner towards minimizing the image similarity criterion between the source and the target image while guaranteeing the expected diffeomorphic nature of the deformation field.
- **Computational Anatomy** [*Nikos Paragios*]
  In [16] we introduced a novel approach for detecting the presence of white matter lesions in periventricular areas of the brain using manifold-constrained embeddings. The proposed method uses locally linear embedding (LLE) to create ”normality” distributions of the brain where deviations from the manifolds are estimated by calculating geodesic distances along locally linear planes in the embedding. Experiments highlight the need of nonlinear techniques to learn the studied data leading to outstanding detection rates when comparing individuals to a specific pathological pattern.

GENSCALE Team

6. New Results

6.1. Next Generation Sequencing

Participants: Alexan Andrieux, Rayan Chikhi, Liviu Ciortuz, Dominique Lavenier, Fabrice Legeai, Claire Lemaitre, Nicolas Maillet, Pierre Peterlongo, Erwann Scaon, Raluca Uricaru.

- **Ultra-low memory data structure for de novo genome assembly** : We propose a new encoding of the de Bruijn graph, which occupies an order of magnitude less space than current representations. The encoding is based on a Bloom filter, with an additional structure to remove critical false positives. [24]
- **Transcriptomic variant detection** : We developped a new method, called kissplice, that calls splicing variant events from sets of RNA-seq NGS reads. It constructs the de-Bruijn graph from the reads and then detects in this graph all patterns corresponding to alternative splicing events. [21]
- **Targeted assembly of NGS data**: The method is based on an iterative targeted assembler which processes large datasets of reads on commodity hardware. Basically, it checks for the presence of given regions of interest in the reads and reconstructs their neighborhood, either as a plain sequence (consensus) or as a graph (full sequence structure). [20]
- **Mapping reads on a graph:** We developped a strategy for directly mapping sequences on bi-directed de-Bruijn graphs. Based on a seed-and-extend algorithm it can be applied on large datasets.[31]
- **Pea aphid genomics and evolution.** Using some of the softwares developped by Genscale, genomic variants and expression data of the pea aphid were analysed, revealing candidate regions involved in the adaptation to host plant, and genes involved in the reproduction mode, either with differential expression patterns or particular patterns of evolutionary rates in other aphid species. [11], [12], [19]

6.2. Protein structures

Participants: Rumen Andonov, Guillaume Chapuis, Dominique Lavenier, Mathilde Le Boudic-Jamin, Antonio Mucherino.

- **Comparison of pairwise protein structure alignments.** The method provides either optimal, top-scoring alignments or heuristic alignments with quality guarantee for some inter-residue distance-based measures. Alignments are compared using a number of quality measures and intuitive visualizations. The methodology brings new insight into the structural relationship of the protein pairs and is a valuable tool for studying structural similarities. [23]
- **Alignment graph.** This object is the main input to find similarities between biomolecules (ARN, proteins). This kind of graph has to model physical and/or chemical properties of the biomolecules and need to take into account constraints dictated by the type of applications (3D comparison, docking, etc.). Our research aims to provide a strategy to automate the building of alignment graphs. A prototype software, called MAGE, is currently under test to validate our approach.
- **Mathematical model and exact algorithm for optimally aligning protein structures.** The algorithm proposes for the first time, to evaluate the popular DALI heuristic in sound mathematical terms. The results indicate that DALI usually computes optimal or close to optimal alignments. However, we detect a subset of small proteins for which DALI fails to generate any significant alignment, although such alignments do exist [22].

- **Modeling the protein flexibility by distance geometry.** We suggest a strategy for modeling protein flexibility that is based on the discretization of the space of possible molecular conformations for a protein. The same discretization process was previously employed for discretizing Molecular Distance Geometry Problems (MDGPs) [30].
- **NMR problems.** We introduce formally the Discretizable Molecular Distance Geometry Problem (DMDGP) for solving the 3D structure of a protein based on Nuclear Magnetic Resonance data together with an algorithm, which we named the "Branch & Prune" (BP), for the solution of DMDGPs [16].We also provide surveys on these recent works about DMDGPs [15], [27].
- **Improvements and variants of the DMDGP**. We exploit symmetries in DMGP trees. We consider similar or related problems (re-ordering of the vertices, relaxing vertices consecutivity assumtion, including side chains and finding low energy homopolymer conformations). Parallelism has also been investigated. [17], [14], [18], [28], [26], [29]

6.3. High Throughput Sequence Analysis

Participants: Rayan Chikhi, Erwan Drezen, Dominique Lavenier, Claire Lemaitre, Nicolas Maillet, Pierre Peterlongo.

- **Comparing metagenomes.** This research aims to define new ways of comparing billions of sequences generated by NGS sequencers. Standard techniques don't scale with such volume of data, both in terms of memory fingerprint and execution time. We have successfully tested a new method based on probabilistic data structures (Bloom filter) allowing large sets of sequences to be indexed in a short time on standard computers. [25]
- **Bank-to-bank comparison.** In cooperation with the Korilog company we improve the PLAST technology developed for bank-to-bank sequence similarity search. Structuration of the index has been revisited to reduce the memory fingerprint and the execution time. The Korilog company has successfully integrated this improvments sofware component in its own software and has just began its promotion with promising responses from several potential clients.[Korilog promotion]

6.4. HPC and Parallelism

Participants: Rumen Andonov, Guillaume Chapuis, Charles Deltel, Dominique Lavenier, Fabrice Legeai, François Moreews.

- **High performance pipelines for annotation**. We participed to effort of URGI (INRA Versailles) to set up TriAnnot, a modular architecture allowing for the annotation of genomes. The TriAnnot pipeline is parallelized on a 712 CPU computing cluster that can run a 1-Gb sequence annotation in less than 5 days. [13]
- **Bioinformatics Workflows.** SLICEE is an environment to capture and parallelize time-consuming bioinformatics applications on grid or cloud platforms. In 2012, a web interface has been designed to interactively draw and run workflows from standard browsers ([workflow portal]). Several workflows used in the BioWIC ANR project have been successfully tested on this platform (http://biowic.inria.fr/)
- **Parallelization of a pseudo-clique solver.** Following such solvers as DAST and A_purva, we develop a pseudo-clique solver for alignment graphs. Looking for pseudo-cliques allows us to relax some of the constraints that are inherent to clique finding and thus maintain polynomial run times. We focus on defining a parallel algorithm and developping an implementation that benefits from multiple levels of parallelism: fine grain parallelism (bit-level parallelism, SSE instructions) and coarse grain parallelism (mutli-core parallelism). Intended applications range from protein local similarity search to protein surface similarity search or even docking.

IBIS Project-Team

5. New Results

5.1. Analysis of gene regulatory networks by means of piecewise-linear (PL) models

GENETIC NETWORK ANALYZER (GNA) is a tool for the qualitative modeling and simulation of the dynamics of gene regulatory networks by means of PL models, as described in Section 4.1 . GNA has been integrated with the other bioinformatics tools distributed by Genostar (http://www.genostar.com/). Version 8.4 of GNA was released by IBIS and Genostar this year. This version is an update of version 8.0, deposited at the Agence pour la Protection des Programmes (APP). Some bugs have been corrected in the new version and the program has been adapted to the latest versions of Java and the software platform of Genostar. A book chapter describing the current version of GNA has been published in a volume on the modeling of bacterial molecular networks [13]. The chapter is a tutorial illustrating the practical use of recent functionalities of GNA like the network editor and the formal verification module by means of an example network in *E. coli*.

The predictions obtained with the help of GNA are purely qualitative, describing the dynamics of the network by means of a state transition graph. While a qualitative analysis is appropriate for certain problems, the absence of precise quantitative predictions may not be desirable in others, such as the analysis of a limit cycle or the design of a controller for a synthetic network. The quantitative study of PL models of gene regulatory networks is hindered by the fact that the step functions describing the logic of regulatory interactions lead to discontinuities in the right-hand side of the PL models (Section 3.2). This has motivated extensions of the PL models based on differential inclusions and Filippov solutions. As of now, no numerical simulation tool for the simulation of these Filippov extensions is available. In collaboration with the BIPOP project-team, we have shown how tools developed for the simulation of nonsmooth mechanical, electrical and control systems can be adapted for this purpose. A paper describing these results is being prepared for submission.

5.2. Experimental mapping of gene regulatory networks in bacteria

The use of fluorescent and luminescent reporter genes allows real-time monitoring of gene expression, both at the level of individual cells and cell populations (Section 3.3). In order to fully exploit this technology, we need methods to rapidly construct reporter genes, both on plasmids and on the chromosome, mathematical models to infer biologically relevant quantities from the primary data, and computer tools to achieve this in an efficient and user-friendly manner. For instance, in a typical microplate experiment, 96 cultures are followed in parallel, over several hours, resulting in 10,000-100,000 measurements of absorbance and fluorescence and luminescence intensities. Over the past few years, we put into place an experimental platform and data analysis software, notably the WELLREADER program (Section 4.2), to allow biologists to make the most of the information contained in reporter gene expression data. Several improvements of the platform for measuring gene expression are the subject of ongoing work, including a novel method for efficiently cloning reporter gene constructions on the chromosome of *E. coli*.

These tools have been used in a series of studies directed at the experimental mapping of gene regulatory networks in *E. coli*. One example, carried out in the framework of the PhD thesis of Guillaume Baptist, is the development of a new screening methodology for identifying all genes that control the expression of a target gene through genetic or metabolic interactions. The screen combines mutant libraries with luciferase reporter constructs. Instead of a static picture of gene expression, this method allows dynamical monitoring in different environmental conditions. Mutants with interesting phenotypes can thus be selected based on multiple criteria, and the expression dynamics of the target gene can be extensively characterized. The method has been applied to the identification of the direct and indirect regulators of the gene *acs* in *Escherichia coli*. We confirmed known genetic regulators of the gene and identified new regulatory influences, many of which involve metabolic intermediates or metabolic sensing. An analysis of mutants involved in glycolysis and glucose transport demonstrates that the classical model of catabolite repression in *E. coli* needs to be amended. A paper describing the above work is currently under revision.

Other examples of on-going work are the analysis of the network involved in motility and sessility and the modulation of the RpoS regulon in *E. coli* by Stephan Lacour, the analysis of the regulation of cAMP levels in the bacterial cell by Claire Villiers, and the analysis of various aspects of the regulation of carbon metabolism by Valentin Zulkower and Stéphane Pinhal.

5.3. Analysis of metabolic coupling in gene regulatory networks

The regulation of gene expression is tightly interwoven with metabolism and signal transduction. A realistic view of genetic regulatory networks should therefore not only include direct interactions resulting from transcription regulation, but also indirect regulatory interactions mediated by metabolic effectors and signaling molecules. We coined the term metabolic coupling to denote these indirect interactions mediated by metabolism. Ignoring metabolic coupling during the analysis of the network dynamics may lead crucial feedback loops to be missed.

In previous work, published in *PLoS Computational Biology* in 2010, we showed how indirect interactions arising from metabolic coupling can be derived from a model of the underlying biochemical reaction network. We applied this approach to the carbon assimilation network in *Escherichia coli* investigating how the structural properties of the network are modified by the inclusion of metabolic interactions. Our results showed that the derived gene regulatory network is densely connected, contrary to what is usually assumed. Moreover, we found that the signs of the indirect interactions are largely fixed by the direction of metabolic fluxes, independently of specific parameter values and rate laws, and that a change in flux direction may invert the sign of indirect interactions. This leads to a feedback structure that is at the same time robust to changes in the kinetic properties of enzymes and that has the flexibility to accommodate radical changes in the environment.

It remains an open question, however, to which extent the indirect interactions induced by metabolic coupling affect the dynamics of the system. This is a key issue for understanding the relative contributions of the regulation of gene expression and metabolism during the adaptation of the cell to changes in its environment. In collaboration with Valentina Baldazzi, formerly post-doctoral fellow in IBIS and now research scientist at INRA (Avignon), we have carried out a dynamic analysis by developing a qualitative PL model of the gene regulatory network, including both the direct and indirect interactions.

In order to obtain a clearer view of the dynamic role of metabolic coupling in the adaptation of gene expression, we developed several qualitative models corresponding to a network topology including all, some, or none of the indirect interactions. The dynamical properties of the models were analyzed and compared with available experimental data using the computer tool GNA (Section 4.1). In particular, we compared the steady-state concentrations of enzymes and transcription regulators during growth on glucose and acetate, as well as the dynamic response of gene expression to the exhaustion of glucose and the subsequent assimilation of acetate. We find significant differences between the dynamics of the system in the absence and presence of metabolic coupling. This confirms that indirect interactions are essential for correctly reproducing the observed adaptation of gene expression to a change in carbon source. Our work thus underlines the importance of metabolic coupling in gene regulatory networks, and shows that such indirect interactions cannot be neglected when studying the adaptation of an organism to changes in its environment. A paper describing these results has been published in the *Journal of Theoretical Biology* [5]. Another publication, reviewing the applicability of these and other ideas for multi-scale modeling in plants, has appeared in *Trends in Plant Science* [4].

5.4. Parameter estimation for kinetic models of carbon metabolism in bacteria

Kinetic models capture the dynamics of the large and complex networks of biochemical reactions that endow bacteria with the capacity to adapt their functioning to changes in the environment. In comparison with the qualitative PL models described in Sections 5.1 and 5.3 , these more general classes of ODE models are intended to provide a quantitative description of the network dynamics, both on the genetic and metabolic level. New experimental techniques have led to the accumulation of large amounts of data, such as time-course measurements of metabolite, mRNA and protein concentrations and measurements of metabolic fluxes under different growth conditions. However, the estimation of parameter values in the kinetic models from these data remains particularly challenging in biology, mostly because of incomplete knowledge of the molecular mechanisms, noisy, indirect, heterogeneous, and partial observations, and the large size of the systems, with dynamics on different time-scales. We have addressed parameter estimation in the context of the analysis of the interactions between metabolism and gene expression in carbon metabolism in *E. coli*.

In collaboration with Matteo Brilli and Daniel Kahn (INRA and Université Claude Bernard in Lyon), we previously developed an approximate model of central metabolism of *E. coli*, as described in an article published in *Bioinformatics* in 2011. The model was based on the use of so-called linlog functions to approximately describe the rates of enzymatic reactions. More precisely, linlog models define reactions rates as proportional to both the enzyme concentrations and a linear combination of the logarithms of metabolite concentrations. The estimation of parameters in the linlog model from metabolomics, transcriptome, proteomics data sets required the development of a new approach, adapted to the occurrence of numerous missing values in the data sets. When applied to the above-mentioned linlog model, exploiting a high-throughput dataset published in the literature, we were able to obtain reasonable estimates of the 100 parameters.

The results of the above application also revealed the fundamental role played by the identifiability of the model parameters, an issue often overlooked in systems biology. This prompted us for a thorough investigation of the concepts of structural identifiability (in presence of perfect, idealized data), practical identifiability (in presence of noisy and limited amounts of data), and the relations among the two. In addition, we looked into the implications of this analysis for the reduction of nonidentifable to identifiable models. While having a solid mathematical basis, the study was tailored to the actual experimental practice, and resulted in a practical model reduction method that improves upon our previous approach in case of large measurement noise. This study, and the results from its application to both *in-silico* case studies and state-of-the-art datasets, were reported in a paper that has been accepted for publication in the *Journal of Mathematical Biology* [6] (see also [11] for a short version with preliminary results).

A second line of work is based on the use of classical kinetic models that are, in comparison with the above-mentioned linlog models, much reduced in scope (the focus is on the metabolic and genetic regulation of the glycolysis pathway) and granularity (individual reactions are lumped together). The models, developed by Delphine Ropers, have been calibrated using experimental data from the experimental part of the IBIS group for the gene expression measurements and the group of Jean-Charles Portais at INSA in Toulouse for the measurements of metabolism. The model with the estimated parameter values is currently being tested and used to understand some key mechanisms in the adaptation of *E. coli* to the exhaustion of glucose. The PhD thesis of Manon Morin, which started at the end of this year in the framework of a collaboration supported by a Contrat Jeune Scientifique INRA-Inria, will further develop these research directions.

5.5. Structural identification of gene regulatory networks

In general, structural identification of genetic regulatory networks involves fitting appropriate network structures and parameters to the data. While modern measurement techniques such as reporter gene systems provide data of ever-increasing quality, the problem remains challenging because exploring all possible network structures in the search of the best fitting model is prohibitive.

In order to address the structural identification problem, Eugenio Cinquemani developed in collaboration with the Automatic Control Lab at ETH Zürich (Switzerland) and the Computer Engineering & Systems Science Department of the University of Pavia (Italy), an ODE modelling framework based on so-called unate-like functions, and a method that exploits monotonicity properties of these functions to effectively prune models that are incompatible with the data from the family of all unate-like modelling alternatives. This model invalidation step is based on simple preprocessing of time-course protein concentration and synthesis rate profiles, assumed available, and allows one to reduce the search of the best fitting model to a small subset of viable model structures.

The method, first published in *Bioinformatics* in 2010 and demonstrated on real data from the synthetic network IRMA, allows one to integrate *a-priori* knowledge on the expected network dynamics in a natural way. Leveraging on this, in the context of the same international collaboration, the method has been further developed in particular by considering relevant subclasses of the family of unate-like models that also enjoy certain quasi-convexity properties. For this restricted class, combined use of monotonicity and quasi-convexity properties allows one to ameliorate the model invalidation step, *i.e.* retain even fewer viable model structures based on affordable data preprocessing. These developments have been presented and demonstrated *in silico* in a paper published in the 2012 special issue on System Identification for Biological Systems of the *International Journal of Robust and Nonlinear Control* [9] .

We are currently applying the above methods to actual, known or partially unknown, networks. In the framework of the PhD thesis of Diana Stefan, the network inference method has been applied to gene expression data from the network regulating motility of *E. coli*. First encouraging results have suggested further experimental and computational investigations that are currently in progress.

5.6. Stochastic modeling and identification of gene regulatory networks in bacteria

At the single-cell level, the processes that govern gene expression are often better described by stochastic models. Modern techniques for the real-time monitoring of gene expression in single cells enable one to apply stochastic modelling to study the origins and consequences of random noise in response to various environmental stresses, and the emergence of phenotypic variability. The potential impact of single-cell stochastic analysis and modelling is tremendous, ranging from a better comprehension of the biochemical regulatory mechanisms underlying life, to the development of new strategies for the control of cell populations and even of single cells. General modeling paradigms, such as the Chemical Master Equation, exist for the description of stochastic dynamics at the single-cell level. However, due to the complexity of the interactions, current studies have often preferred to focus on specific cases of interest by *ad-hoc* modeling and analysis. In addition, theoretical and practical challenges inherent in the inference of stochastic models from biological experimental data have limited the development of general identification approaches.

In view of the potential and the relevance of the subject, one research line of IBIS is dedicated to the probabilistic modeling of the dynamics of gene regulatory networks at the level of individual cells. Our activity is centered around two main challenges. On the one hand, we address the problem of developing methods for fitting unknown network parameters of stochastic models to experimental data. As a reference case study we consider the network regulating the inset of the arabinose uptake process in *E. coli* upon depletion of glucose in the growth medium. For this system, Eugenio Cinquemani and Michel Page are developing and implementing methods for the inference of unknown parameters from fluorescence microscopy data. On the other hand, we investigate several alternative modelling approaches in an attempt to determine their relevance to different systems and application scenarios. This activity is being developed in collaboration with Gregory Batt (CONTRAINTES, Inria Paris-Rocquencourt), Giancarlo Ferrari-Trecate (University of Pavia, Italy), and Alfonso Carta (COMORE, Inria Sophia-Antipolis - Méditerranée). First results connected to control applications on real and simulated data have been submitted for presentation at the European Control Conference to be held in 2013. Finally, further ongoing work concerns the study of noise propagation in gene regulatory networks, in collaboration with Irina Mihalcescu (Université Joseph Fourier), and the analysis of data from Fluorescence Recovery After Photobleaching (FRAP) experiments, in collaboration with Marianna Rapsomaniki and Zoi Lygerou (University of Patras, Greece) and John Lygeros (ETH Zürich, Switzerland).

5.7. Control of regulatory networks in bacteria

While systems biology is primarily concerned with natural systems shaped by evolution, synthetic biology opens up a new generation of fundamental research by trying to redesign natural systems or create novel systems from scratch. Mathematical modeling and analysis are essential components of synthetic biology, as they help understanding the consequences of (changes in) the network of interactions on the dynamical behavior of the system. More specifically, a model can be a powerful tool for the control and regulation of the system towards a desired goal.

Within the projects ColAge and GeMCo (Section 7.2), we attempt to control one of the fundamental physiological properties of bacterial cells, their growth rate. In particular, in order to control the growth rate, we propose to focus on the gene expression machinery of *E. coli*, whose activity is controlled by a complex regulatory network with many components and intertwined feedback loops. Delphine Ropers is developing models of the gene expression machinery and Jérome Izard, in the context of his PhD thesis, is rewiring part of the network to enable control of the network dynamics. The results on these projects are currently being prepared for publication.

5.8. Shared control of gene expression by global physiological effects and specific regulators

Gene expression is controlled by the joint effect of (i) the global physiological state of the cell, in particular the activity of the gene expression machinery, and (ii) DNA-binding transcription factors and other specific regulators. While many studies have focused on networks of transcription factors, the analysis of the relative contributions of both transcription factors and global effects of the physiological state has received relatively little attention thus far.

In the framework of the PhD thesis of Sara Berthoumieux, we have developed a model-based approach to distinguish between these two effects using time-resolved measurements of promoter activities. We have demonstrated the strength of the approach by analyzing a circuit involved in the regulation of carbon metabolism in *E. coli*, consisting of two pleiotropic regulators of the cell (Crp and Fis), the gene *acs* encoding the enzyme acetyl-CoA synthetase (Acs), and the signaling metabolite cyclic AMP (cAMP) which activates Crp. *acs* is strongly expressed in the absence of glucose and is thus an excellent indicator of the transcriptional response of carbon metabolism to a growth-phase transition.

Our results show that the transcriptional response of the network is controlled by the physiological state of the cell and the signalling metabolite cAMP. The (surprising) absence of a strong regulatory effect of transcription factors suggests that they are not the main coordinators of gene expression changes during growth transitions, but rather that they complement the effect of global physiological control mechanisms. This change of perspective has important consequences for the interpretation of transcriptome data and the design of biological networks in biotechnology and synthetic biology. An article presenting the above results has been accepted for *Molecular Systems Biology* [7].

MACS Project-Team

6. New Results

6.1. Asymptotic and multiscale modeling in biomechanics

6.1.1. Detailed validations of muscle model

Participants: Matthieu Caruel, Dominique Chapelle, Alexandre Imperiale, Philippe Moireau.

Until recently we had only considered simple isotropic passive laws of Mooney-Rivlin type in our muscle model, albeit with an overall behavior already highly non-isotropic due to the fiber-oriented active component. We have now implemented and calibrated a new visco-hyperelastic passive law of exponential and orthotropic type for the hyperelastic part, in better agreement with the models and data generally found in the literature. It should be noted that most experimental data available concern the passive behavior only, indeed. In addition, we implemented a new conservative numerical scheme for the time discretization of the contractile variables. Moreover an original boundary condition of contact type has been successfully applied on several detailed cardiac geometries to represent the interactions between the epicardium, pericardium and the surrounding structures.

Major advances in the understanding of heart contraction cycle can be achieved by testing papillary muscle preparations *in vitro*. Single papillary muscles have an essentially one-dimensional structure suitable for uniaxial mechanical testing, and therefore represent the simplest setup to test the robustness of a model of heart contraction against a vast set of experimental results available in the literature. In collaboration with Y. Lecarpentier (Institut du Coeur, Pitié-Salpêtrière Hospital, Paris) and R. Chabiniok (King's College London), we have further refined and calibrated the muscle mechanical model in order to quantitatively reproduce experimental data from rat cardiomyocytes. These results include the static stress-strain constitutive relation, kinetic response to isotonic loadings, and force-velocity relation see Fig.1 .

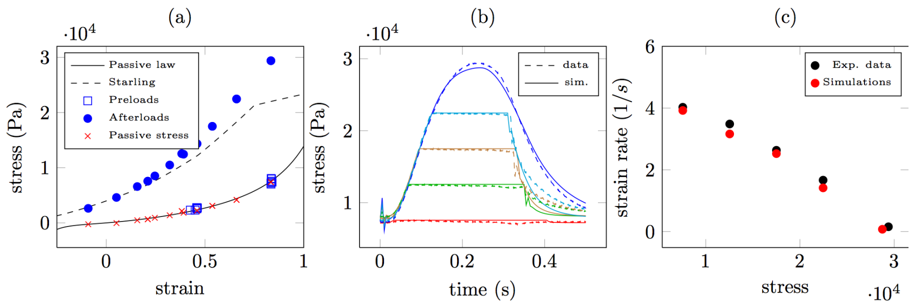


Figure 1. Results of the model compared with experimental data

6.1.2. Multi-scale mechanics of muscle contraction

Participant: Matthieu Caruel.

Muscles are an active tissue material capable of producing force. At the microscale, force is the result of complex interactions between two types of proteins, namely, actin and myosin, which work coherently in very large assemblies ($\sim 10^9$). The passive mechanical response of so-called striated muscles at fast time scales is dominated by long range interactions inducing cooperative behavior without breaking the detailed balance. This leads to such unusual material properties as negative equilibrium stiffness and drastically different behavior in force and displacement controlled loading conditions. Analysing experimental data strongly suggests that muscles are finely tuned to perform close to a critical point (see Fig.2). This work in collaboration with Jean-Marc Allain and Lev Truskinovsky (LMS, Ecole Polytechnique) is the subject of a paper submitted to Physical Review Letters (see [22]).

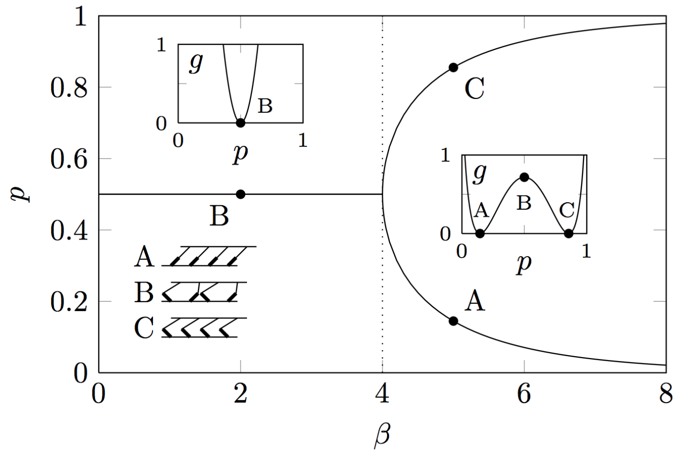


Figure 2. Bifurcation diagram of a model of coupled molecular motors. p is the fraction of motors in the stress generating configuration (post-power-stroke). β is a non dimensional parameter representing the intensity of thermal fluctuations ($\beta \to 0$ represents infinitely strong thermal forces). For $\beta < 4$, the system lives in a mixed configuration (B): the free energy is convex with a minimum at $p = 1/2$ (see the left inset showing the energy landscape g). For $\beta > 4$ the system is organized in two distinct populations (A and C) corresponding to the 2 minima of a non-convex energy landscape (see the right inset). One population is mainly pre-power-stroke (A) while the other is post-power-stroke (C). This is a signature of mechanical cooperativity.

6.1.3. Asymptotic analysis applied to cardiac electrophysiology modeling

Participants: Dominique Chapelle, Annabelle Collin, Jean-Frédéric Gerbeau [(REO team)].

Computational electrophysiology is a very active field with tremendous potential in medical applications, albeit leads to highly intensive simulations. We here propose a surface-based electrophysiology formulation, motivated by the modeling of thin structures such as cardiac atria, which greatly reduces the size of the computational models. Moreover, our model is specifically devised to retain the key features associated with the anisotropy in the diffusion effects induced by the fiber architecture, with rapid variations across the thickness which cannot be adequately represented by naive averaging strategies. Our proposed model relies on a detailed asymptotic analysis in which we identify a limit model and establish strong convergence results. We also provide detailed numerical assessments which confirm an excellent accuracy of the surface-based model – compared with the reference 3D model – including in the representation of a complex phenomenon, namely, spiral waves, see Figure 3 . This work was submitted for publication in "M3AS: Mathematical Models and Methods in Applied Sciences".

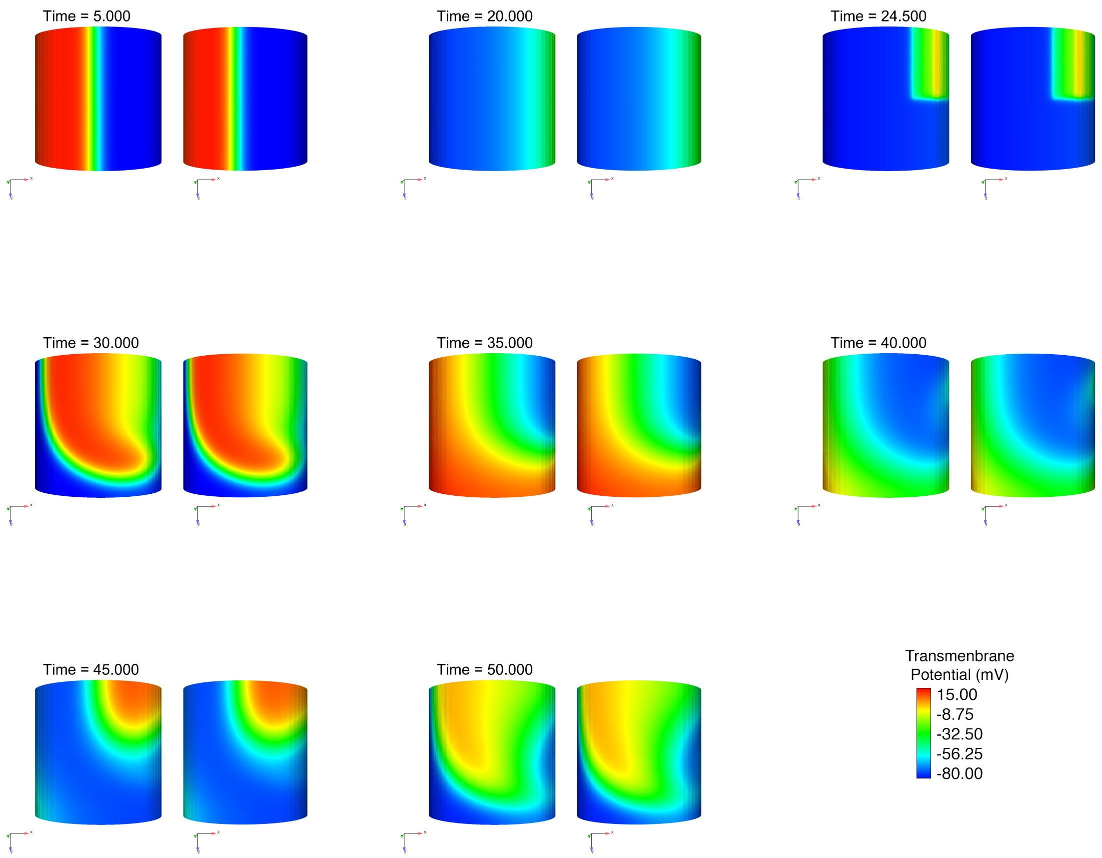


Figure 3. Spiral wave on cylinder – Comparison of asymptotic surface model (left), 3D model (center) and naive 2D model (right) on the midsurface at 8 consecutive times

6.1.4. Cardiac atria electrophysiology surface-based modeling and assessment of physiological simulations

Participants: Dominique Chapelle, Annabelle Collin, Jean-Frédéric Gerbeau [(REO team)].

We aim at validating the 2D (namely, surface-based) electrophysiology model designed for thin cardiac structures with strongly heterogeneous anisotropy, presented in Paragraph 6.1.3 with a real model of the atria. We produced a surface mesh representing the mid-surface of the two atria. We used the bibliography to identify and prescribe the fibers directions at the endocardium and epicardium. Figure 4 displays the simulation results obtained with the surface-based model.

6.1.5. Strong convergence results in the asymptotic behavior of the 3D-shell model

Participants: Dominique Chapelle, Annabelle Collin.

The objective of this work is to establish the strong convergence for the asymptotic analysis of the so-called 3D-shell model presented in [2]. We apply similar methods to those used in the work on "Asymptotic analysis applied to cardiac electrophysiology modeling".

6.2. Estimation in biomechanics

6.2.1. Exponential convergence of an observer based on partial field measurements for the wave equation

Participants: Dominique Chapelle, Philippe Moireau.

We analyze an observer strategy based on partial—that is, in a subdomain—measurements of the solution of a wave equation, in order to compensate for uncertain initial conditions. We prove the exponential convergence of this observer under a nonstandard observability condition, whereas using measurements of the time derivative of the solution would lead to a standard observability condition arising in stabilization and exact controllability. Nevertheless, we directly relate our specific observability condition to the classical geometric control condition. Finally, we provide some numerical illustrations of the effectiveness of the approach. This work in collaboration with M. de Buhan (Univ. Paris V) and N. Cîndea (Univ. Clermont-Ferrand) is published in [13].

6.2.2. Sequential identification of boundary support parameters in a fluid-structure vascular model using patient image data

Participants: Dominique Chapelle, Philippe Moireau.

This work [17] is in collaboration with C. Bertoglio and J.-F. Gerbeau (REO team) and N. Xiao, C.A. Figueroa and C.A. Taylor (Stanford University), where we propose a complete methodological chain for the identification of the corresponding boundary support parameters, using patient image data. We consider distance maps of model to image contours as the discrepancy driving the data assimilation approach, which then relies on a combination of (1) state estimation based on the so-called SDF filtering method, designed within the realm of Luenberger observers and well-adapted to handling measurements provided by image sequences, and (2) parameter estimation based on a reduced-order UKF filtering method which has no need for tangent operator computations and features natural parallelism to a high degree. Implementation issues are discussed, and we show that the resulting computational effectiveness of the complete estimation chain is comparable to that of a direct simulation. Furthermore, we demonstrate the use of this framework in a realistic application case involving hemodynamics in the thoracic aorta. The estimation of the boundary support parameters proves successful, in particular in that direct modeling simulations based on the estimated parameters are more accurate than with a previous manual expert calibration. This paves the way for complete patient-specific fluid-structure vascular modeling in which all types of available measurements could be used to estimate additional uncertain parameters of biophysical and clinical relevance.

This work published in BMMB (impact factor 3.192) can be considered as the first trial of data assimilation using real data in hemodynamics.

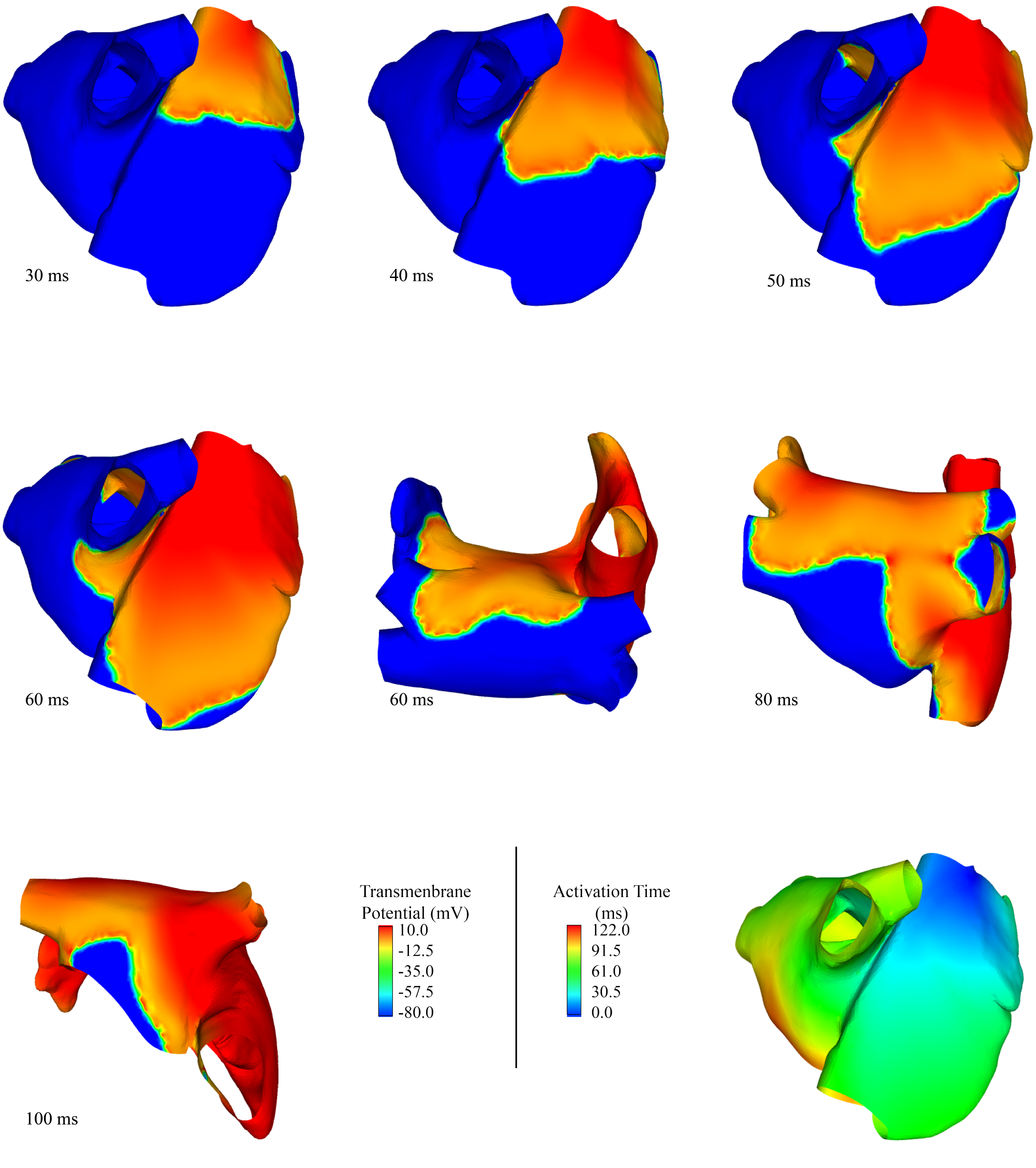


Figure 4. Simulation of atrial depolarization

6.2.3. Filtering strategies using image data

Participants: Alexandre Imperiale, Philippe Moireau, Alexandre Routier.

Some progress has been achieved concerning the Luenberger filtering procedure – also known as nudging – for the cardiovascular system in several directions. We have studied the impact of data interpolation (in time and space) on the method performance (a paper on this subject is being prepared) and, during Alexandre Routier end-of-curriculum internship from INSA Rouen, we have adapted the formalism of currents (inspired by a collaboration with S. Durrleman). This formalism in an elegant way to represent geometric objects (endo- and epicardium surfaces for example) as operators on a test vector space defined on the ambient space. From this key idea the main work was to define a numerical tractable norm on the space of surfaces and derive it with respect to the Lagrangian displacement of the solid domain in order to incorporate such a representation of surfaces into our filtering technique. Among other advantages this new observer requires significantly less prior efforts in terms of image processing.

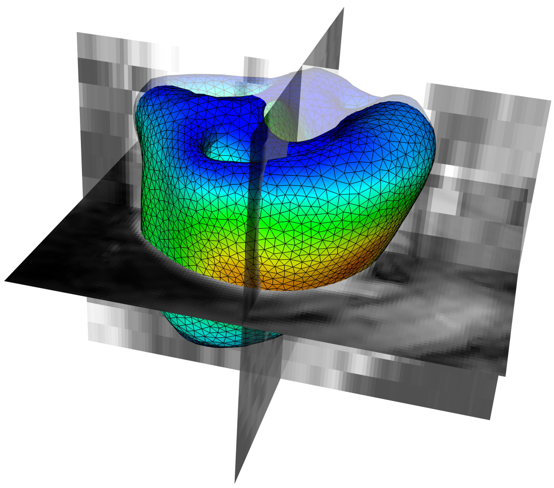


Figure 5. Results of the model compared with experimental data

6.2.4. Formulation of observers for parabolic equations

Participants: Karine Mauffrey, Philippe Moireau.

We are currently working on optimal filtering using observers for a class of evolution PDEs including heat-like equations. As for the optimal control issue,the optimal filtering issue is related to the resolution of a differential Riccati equation. In [25] or [29], the link between the optimal filtering formulations and the derived Riccati equation is done by finite dimension arguments. There exist also other results on the linear quadratic optimal control that are based on infinite dimensional considerations (see, for example, [26] and [28]). A work in progress consists in presenting a direct approach for the optimal filtering issue, using infinite dimension considerations only. Then we should be able to introduce reduced-rank considerations to be able to stabilize only the low frequencies part of the parabolic system, and therefore offer a discretization strategy. This discretization will the be analyzed in details.

6.3. Other topics

6.3.1. Sail modeling

Participants: Dominique Chapelle, Daniele Trimarchi.

This is a collaboration with Marina Vidrascu (REO team) and Stephen Turnock and Dominic Taunton (Southampton University), as part of the recently completed PhD of Daniele Trimarchi. We propose a method of modelling sail type structures which captures the wrinkling behaviour of such structures. The method is validated through experimental and analytical test cases, particularly in terms of wrinkling prediction. An enhanced wrinkling index is proposed as a valuable measure characterizing the global wrinkling development on the deformed structure. The method is based on a pseudo-dynamic finite element procedure involving non-linear MITC shell elements. The major advantage compared to membrane models generally used for this type of analysis is that no ad hoc wrinkling model is required to control the stability of the structure. We demonstrate our approach to analyse the behaviour of various structures with spherical and cylindrical shapes, characteristic of downwind sails over a rather wide range of shape and constitutive parameters. In all cases convergence is reached and the overall flying shape is most adequately represented, which shows that our approach is a most valuable alternative to standard techniques to provide deeper insight into the physical behaviour. Limitations appear only in some very special instances in which local wrinkling-related instabilities are extremely high and would require specific additional treatments, out of the scope of the present study. This work has been published in [20].

6.3.2. PODs for parameter-dependent problems and estimation

Participants: Dominique Chapelle, Philippe Moireau.

This work – submitted to M2AN [24] – is derived from the latest part of Asven Gariah's PhD, jointly supervised by Jacques Sainte-Marie (Bang team) and D. Chapelle, and defended in late 2011. We address the issue of parameter variations in POD approximations of time-dependent problems, without any specific restriction on the form of parameter dependence. Considering a parabolic model problem, we propose a POD construction strategy allowing us to obtain some *a priori* error estimates controlled by the POD remainder – in the construction procedure – and some parameter-wise interpolation errors for the model solutions. We provide a thorough numerical assessment of this strategy with the FitzHugh-Nagumo 1D model. Finally, we give detailed illustrations of the approach in two parameter estimation applications, the first in a variational estimation framework with the FitzHugh-Nagumo model, and the second with a beating heart mechanical model for which we employ a sequential estimation method to characterize model parameters using real image data in a clinical case.

MAGIQUE-3D Project-Team

6. New Results

6.1. Inverse Problems

6.1.1. Reconstruction of an elastic scatterer immersed in a homogeneous fluid

Participants: Hélène Barucq, Rabia Djellouli, Élodie Estecahandy.

The determination of the shape of an obstacle from its effects on known acoustic or electromagnetic waves is an important problem in many technologies such as sonar, radar, geophysical exploration, medical imaging and nondestructive testing. This inverse obstacle problem (IOP) is difficult to solve, especially from a numerical viewpoint, because it is ill-posed and nonlinear. Its investigation requires as a prerequisite the fundamental understanding of the theory for the associated direct scattering problem, and the mastery of the corresponding numerical solution methods.

In this work, we are interested in retrieving the shape of an elastic obstacle from the knowledge of some scattered far-field patterns, and assuming certain characteristics of the surface of the obstacle. The corresponding direct elasto-acoustic scattering problem consists in the scattering of time-harmonic acoustic waves by an elastic obstacle Ω^s embedded in a homogeneous medium Ω^f, that can be formulated as follows:

$$\begin{aligned} \Delta p + (\omega^2/c_f^2)\, p &= 0 && \text{in } \Omega^f \\ \nabla \cdot \sigma(u) + \omega^2 \rho_s\, u &= 0 && \text{in } \Omega^s \\ \omega^2 \rho_f u \cdot n &= \partial p/\partial n + \partial e^{i\,(\omega/c_f)\, x\cdot d}/\partial n && \text{on } \Gamma \\ \sigma(u) n &= -pn - e^{i\,(\omega/c_f)\, x\cdot d} n && \text{on } \Gamma \\ \lim_{r\to+\infty} r\,(\partial p/\partial r - i\,(\omega/c_f)\, p) &= 0 \end{aligned} \tag{30}$$

where p is the fluid pressure in Ω^f whereas u is the displacement field in Ω^s, and $\sigma(u)$ represents the stress tensor of the elastic material.

This boundary value problem has been investigated mathematically and results pertaining to the existence, uniqueness and regularity can be found in [86] and the references therein, among others. We propose a solution methodology based on a regularized Newton-type method for solving the IOP. The proposed method is an extension of the regularized Newton algorithm developed for solving the case where only Helmholtz equation is involved, that is the acoustic case by impenetrable scatterers [79]. The direct elasto-acoustic scattering problem defines an operator $F : \Gamma \to p_\infty$ which maps the boundary Γ of the scatterer Ω^s onto the far-field pattern p_∞. Hence, given one or several measured far-field patterns $\widetilde{p}_\infty(\widehat{x})$, corresponding to one or several given directions d and wavenumbers k, one can formulate IOPs as follows:

$$\text{Find a shape } \Gamma \text{ such that } \quad F(\Gamma)(\widehat{x}) = \widetilde{p}_\infty(\widehat{x}); \quad \widehat{x} \in S^1.$$

We propose a solution methodology based on a regularized Newton-type method to solve this inverse obstacle problem. At each Newton iteration, we solve the forward problem using a finite element solver based on discontinuous Galerkin approximations, and equipped with high-order absorbing boundary conditions. We have first characterized the Fréchet derivatives of the scattered field. They are solution to the same boundary value problem as the direct problem with other transmission conditions. This work has been presented both in FACM11 and in WAVES 2011. A paper has been submitted.

6.1.2. *hp-adaptive inversion of magnetotelluric measurements*

Participants: Hélène Barucq, Julen Alvarez Aramberri, David Pardo.

The magnetotelluric (MT) method is a passive electromagnetic (EM) exploration technique that allows to determine the resistivity distribution in the subsurface of the area of interest on scales varying from few meters to hundreds of kilometers. Commercial uses include hydrocarbon (oil and gas) exploration, geothermal exploration, and mining exploration, as well as hydrocarbon and groundwater monitoring. MT measurements are governed by the electromagnetic phenomena, which can be described by Maxwell's equations. We solve those equations by a goal-oriented hp-adaptivity Finite Element Method (FEM).

In order to estimate the resistivity distribution in the Earth's subsurface, we solve an Inverse Problem. We define a Misfit Function that represents the difference between the measured and computed data for a particular resistivity distribution. By minimizing this misfit function using a gradient based approach with model reduction techniques, and hence solving the inverse problem, we are able to determine the properties of the subsurface materials.

6.2. Modeling

6.2.1. *Implementation of a non-reflecting boundary condition on ellipsoidal boundary*

Participants: Hélène Barucq, Anne-Gaëlle Saint-Guirons, Sébastien Tordeux.

The modeling of wave propagation problems using finite element methods usually requires the truncation of the computational domain around the scatterer of interest. Absorbing boundary condition are classically considered in order to avoid spurious reflections. This year we have implemented and tested an exact condition based on a non local Dirichlet to Neumann operator in the context of the Helmholtz equation posed on an elongated domain.

6.2.2. *Explicit computation of the electrostatic energy for an elliptical charged disc*

Participants: Sophie Laurens, Sébastien Tordeux.

In [32], We have described a method to obtain an explicit expression for the electro- static energy of a charged elliptical infinitely thin disc. The charge distribution is assumed to be polynomial. Such explicit values for this energy are fundamen- tal for assessing the accuracy of boundary element method codes. The main tools used are an extension of Copson's method and a diagonalization, given by Leppington and Levine, of the single-layer potential operator associated with the electrostatic potential created by a distribution of charges on an elliptical disc.

6.2.3. *A new modified equation approach for solving the wave equation*

Participants: Cyril Agut, Hélène Barucq, Henri Calandra, Julien Diaz, Florent Ventimiglia.

The new method involving p-harmonic operator described in section 3.2 has been presented in [17]. We have proved the convergence of the scheme and its stability under a CFL condition. Numerical results in one, two and three-dimensional configurations show that this CFL condition is slightly greater than the CFL condition of the second-order Leap-Frog scheme.

In the framework of the PhD thesis of Florent Ventimiglia, we are now considering the extension of this technique to the first order formulation of the acoustic and elastodynamic equations. A numerical analysis of performance in 1D indicates that, for a given accuracy, this method requires less storage than the High-Order ADER Schemes for and similar computational costs. We are now implementing this algorithm in 3D in order to confirm this analysis and to assess its performance in an RTM framework on realistic configurations.

6.2.4. *Stability Analysis of an Interior Penalty Discontinuous Galerkin Method for the Wave equation*

Participants: Cyril Agut, Hélène Barucq, Julien Diaz.

The Interior Penalty Discontinuous Galerkin Method [72], [69], [83] we use in the IPDGFEM code requires the introduction of a penalty parameter. Except for regular quadrilateral or cubic meshes, the optimal value of this parameter is not explicitely known. Moreover, the condition number of the resulting stiffness matrix is an increasing function of this parameter, but the precise behaviour has not been explicited neither. We have carried out a theoretical and numerical study of the pnealization parameter and of the CFL condition for quadrilateral and cubic meshes, this results have been presented in a paper accepted in M2AN [16]

6.2.5. Higher Order Absorbing Boundary Conditions for the Wave Equation

Participants: Hélène Barucq, Juliette Chabassier, Julien Diaz.

The numerical simulation of wave propagation is generally performed by truncating the propagation medium and the team works on new ABCs, trying to improve the performance of existing conditions. Following the analysis performed in [23], we have considered the issue of constructing high-order ABCs for the Helmholtz equation. Now, to derive conditions of order greater than two is really technical. In addition, when the coefficients representing the geological properties of the medium are not regular, the method of construction of ABCs is not completely justified. That is why we turned to the construction of conditions that take into account all the characteristics of the diffraction phenomenon and not only waves that propagate like in the case of standrad ABCs. This is what we call enriched ABCs. A research report is being written, an article should be submitted in 2013. During 2012, a publication for the acoustic wave equation has been accepted in M3AS [23] and a second one has been submitted.

6.2.6. Multiperforated plates in linear acoustics

Participants: Abderrahmane Bendali, M'Barek Fares, Sophie Laurens, Estelle Piot, Sébastien Tordeux.

Acoustic engineers use approximate heuristic models to deal with multiperforated plates in liners and in combustion chambers of turbo-engines. These models were suffering from a lack of mathematical justifications and were consequently difficult to improve. Performing an asymptotic analysis (the small parameter is the radius of the perforations), we have justified these models and proposed some improvement. Our theoretical results have been compared to numerical simulations performed at CERFACS (M'Barek Fares) and to acoustical experiments realized at ONERA (Estelle Piot). Two papers have been published in 2012 [27], [30].

6.2.7. Performance Assessment of IPDG for the solution of an elasto-acoustic scattering problem

Participants: Hélène Barucq, Rabia Djellouli, Élodie Estecahandy.

We present a solution methodology for the direct elasto-acoustic scattering problem that falls in the category of Discontinuous Galerkin methods. The method distinguishes itself from the existing methods by combining high-order Discontinuous Galerkin approximations, local stabilizations for the coupled problem and the use of curved element edges on the boundaries. We present some numerical results that illustrate the salient features and highlight the performance of the proposed solution methodology on the resonance phenomenon existing in the elastic scatterer for simple geometries such as circles. Moreover, the designed method ensures a convergence order with a gain of two order of magnitude compared to polygonal boundaries, and a potential to address both mid- and high-frequency regimes. These results have been presented to ECCOMAS 2012 [44] and to two workshops [42] [43].

6.2.8. Operator Based Upscaling for Discontinuous Galerkin Methods

Participants: Hélène Barucq, Théophile Chaumont, Julien Diaz, Christian Gout, Victor Péron.

Scientists and engineers generally tackle problems that include multiscale effects and that are thus difficult to solve numerically. The main difficulty is to capture both the fine and the coarse scales to get an accurate numerical solution. Indeed, the computations are generally performed by using numerical schemes based on grids. But the stability and thus the accuracy of the numerical method depends on the size of the grid which must be refined drastically in the case of very fine scales. That implies huge computational costs and in particular the limitations of the memory capacity are often reached. It is thus necessary to use numerical methods that are able to capture the fine scale effects with computations on coarse meshes. Operator-based upscaling is one of them and we present in [22] a first attempt to adapt that technique to a Discontinuous Galerkin Method (DGM). We consider the Laplace problem as a benchmark and we compare the performance of the resulting numerical scheme with the classical one using Lagrange finite elements. The comparison involves both an accuracy analysis and a complexity calculus. This work shows that there is an interest of combining DGM with upscaling.

6.2.9. Asymptotic Modeling for Elasto-Acoustics

Participants: Julien Diaz, Victor Péron.

We present in [65] equivalent conditions and asymptotic models for the diffraction problem of elasto-acoustic waves in a solid medium surrounded by a thin layer of fluid medium. This problem is well suited for the notion of equivalent conditions : since the thickness of the layer is small with respect to the wavelength, the effect of the fluid medium on the solid is as a first approximation local. We derive and validate equivalent conditions up to the third order for the elastic displacement. These conditions approximate the acoustic waves which propagate in the fluid region. This approach leads us to solve only elastic equations. The construction of equivalent conditions is based on a multiscale expansion in power series of the thickness of the layer for the solution of the transmission problem.

Questions regarding the implementation of the conditions have been addressed carefully. Indeed, the boundary conditions have been integrated without changing the structure of the code Hou10ni.

This work has been presented in four international conferences and Workshops : Aquitaine-Euskadi Workshop on Applied Mathematics; First Russian-French Conference on Mathematical Geophysics, Mathematical Modeling in Continuum Mechanics and Inverse Problems; Workshop HPC-GA; Twelfth International Conference Zaragoza-Pau on Mathematics.

A paper with numerical results for the elasto-acoustic problem with a thin layer and a variable thickness is in preparation.

6.2.10. Asymptotic modeling in electromagnetism

Participants: François Buret, Monique Dauge, Patrick Dular, Laurent Krähenbühl, Victor Péron, Ronan Perrussel, Clair Poignard, Damien Voyer.

The following results rely on a problematic developed in section 3.2 , item **Asymptotic modeling**.

In the paper [28], eddy current problems are addressed in a bidimensional setting where the conducting medium is non-magnetic and has a corner singularity. For any fixed skin depth we show that the flux density is bounded near the corner, unlike the perfect conducting case. Then as the skin depth goes to zero, the first two terms of a multiscale expansion of the magnetic potential are introduced to tackle the magneto-harmonic problem. The heuristics of the method are given and numerical computations illustrate the obtained accuracy.

In a forthcoming paper, we describe the magnetic potential in the vicinity of a corner of a conducting body embedded in a dielectric medium in a bidimensional setting. We make explicit the corner asymptotic expansion for this potential as the distance to the corner goes to zero. This expansion involves singular functions and singular coefficients. We introduce a method for the calculation of the singular functions near the corner and we provide two methods to compute the singular coefficients: the method of moments and the method of quasi-dual singular functions. Estimates for the convergence of both approximate methods are proven. We eventually illustrate the theoretical results with finite element computations. The specific non-standard feature of this problem lies in the structure of its singular functions: They have the form of series whose first terms are harmonic polynomials and further terms are genuine non-smooth functions generated by the piecewise constant zeroth order term of the operator. This work has been presented in the international conference WCCM 2012.

6.2.11. Asymptotic models for penalization methods in porous media

Participants: Gilles Carbou, Victor Péron.

We investigate a Stokes-Brinkman problem with Beavers and Joseph transmission conditions, adapted to a penalization method in porous media. We exhibit a WKB expansion for the solution of the fluid-porous interface problem. The main interest is to derive equivalent models for the penalization method. We explicit the first terms of the WKB expansion for the flow and the pressure in the subdomains. Each asymptotics of the flow writes as a sum of a tangential boundary layer term plus a standard term in the porous region. From the benefits or these boundary layers, we infer a collection of elementary transmission problems satisfied by the standard parts of the asymptotics for the flow and the pressure. As a consequence of the penalization of the Laplacian operator which applies to the flow in the porous media, a degenerate operator of order zero applies to the elementary velocities appears in the porous region. The main difficulty concern the proof of elliptic regularity up to the interface for the solution of each elementary problem, since exotic conditions for the flow and the pressure appears along the interface. Our strategy consists to adapt a proof of elliptic regularity for the solution of a Darcy problem set in homogeneous media and developed by Boyer-Fabrie.

6.2.12. Asymptotic modeling in electromagnetism

Participants: Marc Duruflé, Victor Péron, Clair Poignard.

We investigate asymptotic models for 3D transmission problems in electromagnetism with homogeneous thin layers (uniform thickness). We exhibit Generalized Impedance Boundary Conditions of order 1 when the thin layer is symmetric and non-symmetric with respect to its mean surface. We present also a limit model for a resistive thin layer, and an equivalent model of order 1 for large contrast in conductivities through the thin layer. We write all these models in a general form. Questions regarding the implementation of the conditions have been addressed carefully. Numerical results with the high-order finite element library Montjoie illustrate the accuracy of the asymptotic models. A paper is in preparation.

6.2.13. Absorbing Boundary Conditions for Tilted Transverse Isotropic Elastic Media

Participants: Hélène Barucq, Lionel Boillot, Henri Calandra, Julien Diaz.

The simulation of wave propagation in geophysical media is often performed in domains which are huge compared to the wavelenghts of the problem. It is then necessary to reduce the computational domain to a box. When considering acoustic or elastic isotropic media, this can be done by applying an Absorbing Boundary Condition (ABC) or by adding a Perfectly Matched Layer (PML). However, a realistic representation of the Earth subsurface must include anisotropy and, in particular, the so-called Tilted Transverse Isotropy. Perfectly Matched Layers are known to be unstable for this kind of media and, to the best of our knowledge, no ABC have been proposed yet. We have thus proposed a low-order ABC for TTI media.

This ABC has been constructed for elliptic TTI media, where the slowness curve of the P-Wave is a rotated ellipse. Then, an appropriate change of variable can be applied in order to transform this ellipse into a circle. The main idea consists in imposing the isotropic ABC in the new system of coordinates and to apply the inverse change of variable in order to obtain the elliptic TTI ABC. We have compared numerically the reflections generated by this new ABC in TTI domain to the ones generated by the classical first order ABC in isotropic domains. The results show that the new ABC performs as well as the classical first order one. Moreover, this ABC seems to be also well-suited to non elliptic TTI media. These results have been presented at the Congrès Français d'Acoustique [35], at two workshops [39], [45].

6.2.14. Efficient solution methodology based on a local wave tracking strategy for high-frequency Helmholtz problems.

Participants: Mohamed Amara, Sharang Chaudhry, Julien Diaz, Rabia Djellouli, Steven Fiedler.

We have designed a new and efficient solution methodology for solving high-frequency Helmholtz problems. The proposed method is a least-squares based technique that employs variable bases of plane waves at the element level of the domain partition. A local wave tracking strategy is adopted for the selection of the basis at the regional/element level. More specifically, for each element of the mesh partition, a basis of plane waves is chosen so that one of the plane waves in the basis is oriented in the direction of the propagation of the field inside the considered element. The determination of the direction of the field inside the mesh partition is formulated as a minimization problem. Since the problem is nonlinear, we apply Newton's method to determine the minimum. The computation of Jacobians and Hessians that arise in the iterations of the Newton's method is based on the exact characterization of the Fréchet derivatives of the field with respect to the propagation directions. Such a characterization is crucial for the stability, fast convergence, and computational efficiency of the Newton algorithm. These results are part of the Master thesis of Sharang Chaudhry (student à CSUN) and have been presented to the 6th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS, Vienna, 2012).

6.3. High Performance methods for solving wave equations

Participants: Lionel Boillot, Hélène Barucq, Henri Calandra, Julien Diaz, Emiljana Jorgji, Didier Rémy, Florent Ventimiglia.

We have recently optimized the DG code implemented in the DIVA plateform of Total by reducing the number of communications between each processors. Since this code is based on the first order formulation of the elastodynamic wave equation, we have to compute three velocities and six stresses at each degree of freedom of the mesh. One naive idea consists in communicating these nine values at each time step. On the other hand, the computation of the three velocities does not actually require the knowledge of the six stresses but of three linear combination of these stresses. Similarly, the computation of the stresses requires the knowledge of six linear combinations of the three velocities. The main idea of the optimization consists in computing the three linear combinations of the stresses and to communicate them to the other processors, while the three velocities are communicated before computing the linear computations. Hence the number of communications can be reduced to six at each time step.

This optimization, coupled with the use of Hybrid MPI and OpenMP parallel programming has allowed to prove the scalability of the code up to 512 cores. We are now planning to extend these tests up to 4000 cores.

MAGNOME Project-Team

6. New Results

6.1. Yeast comparative genomics

Participants: Pascal Durrens [correspondant], Tiphaine Martin, David James Sherman.

By using MAGNOME's MAGUS system and YAGA software, we have successfully realized a full annotation and analysis of seven new genomes, provided to the Génolevures Consortium by the CEA–Génoscope (Évry)[15]. Two distant genomes from the *Debaryomycetaceae* and *mitosporic Saccharomycetales* clades of the *Saccharomycetales* were annotated using previously published Génolevures genomes [6], [10], [11] as references (in prep.). A further group of five species, comprised of pathogenic and nonpathogenic species, was analyzed with the goal of identifying virulence determinants [37]. By choosing species that are highly related but which differ in the particular traits that are targeted, in this case pathogenicity, we are able to focus of the few hundred genes related to the trait (in rev.). The approximately 40,000 new genes from these studies were classified into existing Génolevures families as well as branch-specific families.

In collaboration with partners in the ISVV, Bordeaux, we have assembled and analyzed 12 wine starter yeasts, with the goal of understanding genetic determinants of performance (in prep.).

6.2. Assembly, annotation and comparison of bacterial Omics data

Participants: Elisabeth Bon [correspondant], Laetitia Bourgeade, Pascal Durrens, Aurélie Goulielmakis, Tiphaine Martin, David James Sherman.

Oenococcus oeni is part of the natural microflora of wine and related environments, and is the main agent of the malolactic fermentation (MLF), a step of wine making that generally follows alcoholic fermentation (AF) and contributes to wine deacidification, improvement of sensorial properties and microbial stability. The start, duration and achievement of MLF are unpredictable since they depend both on the wine characteristics and on the properties of the *O. oeni* strains. In collaboration with Patrick Lucas's lab of the ISVV Bordeaux that is currently proceeding with genome sequencing, explorative and, and comparative genomics, Elisabeth Bon coordinates our efforts into the OENIKITA project (since 2009), a scale switching challenge including highthrouput exploratory and comparative genomics for oenological bacterial starters, and the development of an online web-collaborative multigenomic comparative platform based on the the Génolevures database architecture and MAGUS / YAGA systems.

OENI-Genomics: In comparative genomics, we investigated gene repertoire and genomic organization conservation through intra- and inter-species genomic comparisons, which clearly show that the *O. oeni* genome is highly plastic and fast-evolving. Results reveal that the optimal adaptation to wine of a strain mostly depends on the presence of key adaptive loops and polymorphic genes. They also point up the role of horizontal gene transfer and mobile genetic elements in O. oeni genome plasticity, and give the first clues of the genetic origin of its oenological aptitudes[3], [14], [29], [33], [35], [36]. As a result of the scaling out challenge, we participated to the assembly and annotation of 19 fully sequenced *O. oeni* genome variants.

KITA-Genomics (E. Bon, D. Sherman): This project that is focused on the sequencing, assembly, exploration and comparison of the *O. kitaharae* genome, has benefited to an international collaboration involving Dr V. Makeev. MAGNOME was involved into the pilot assembly, exploration and comparison of the *O. kitaharae* genome.

Transcriptomic axis (E. Bon, A. Goulielmakis, P. Durrens): Under the supervision of E. Bon, Aurélie Goulielmakis has completed for the ANR DIVOENI a detailed manual annotation of a new reference strain of *O. oeni* and performed comparative transcriptome analysis to identify genes differentially expressed under different culture conditions. We explored and compared how the expression system is solicited when *O. oeni* strains adapted to grow in some niches are placed under stress-exposure conditions. The monitoring of gene expression status between strains, through the definition of a global expression pattern proper to each gene, partially lift the veil on how *O. oeni* genome adapts function to its environment. The weight of genetic background and ecological niche pressure on gene expression flexibility was evaluated, and the *O. oeni* pan-transcriptome architecture characterized. The first guidelines revealed a supra-spatial organization of stress response into activated and repressed larger macro-domains defining functional landmarks and intra-chromosomal territories. Decryption of stress-sensitive gene repertoires promises to be an efficient tool in the conquest of *O. oeni* "domestication" through the identification of molecular markers responsible for different physiological capabilities, and the selection of the best adapted strains [21], [43].

Gene plasticity modelisation (E. Bon, L. Bourgeade): A novel axis of research recently emerged under the initiative of E. Bon (pseudOE project) around the detection, characterization and conservation of pseudogenes populations in *Oenococcus* bacteria. Such topic presents a double interest: phylogenetic at first because it should allow to better estimate the degree of genic/genomic plasticity of these bacteria, and algorithmic then because the pseudogenes are a source of confusion for the automatic prediction of genes. Through a transversal collaboration and a cooperative supervision with the Algorithms for Analysis of Biological Structures Group (P. Ferraro, J. Allali) at LaBRI, Laetitia Bourgeade (PhD, Univ. Bordeaux1) was recruited to develop dedicated methods to improve pseudogenes automatic detection, and therefore gene predictions, and to reconstruct fossil and modern genes evolutionary history [20], [23].

6.3. Big Data in comparative genomics

Participants: David James Sherman [correspondant], Pascal Durrens, Natalia Golenetskaya, Florian Lajus, Tiphaine Martin.

Data growth in comparative genomics presents a significant scaling challenge that requires novel informatic methods. Increase in sequence data is already a challenge, but in addition, the *relations* between the biological objects increase supralinearly (geometrically in the worst case) for every linear increase in sequence data.

MAGNOME's Tsvetok system proposes a highly-scalable distributed approach for data and computation in comparative genomics, targeting projects of the "comparative genomics of related species" type, where a set of genomes is sequenced and analyzed as part of the same process. Tsvetok combines a novel NoSQL storage schema with domain-specific MapReduce algorithms, to efficiently handle the fundamentally data-parallel analyses encountered in comparative genomics. Natalia Golenetskaya with Florian Lajus derived use cases from web site log analyses to identify standard queries, define an appropriate query-oriented storage schema, and map structured values to this schema. This was tested in MAGNOME's dedicated computing cluster.

Natalia Golenetskaya furthermore defined new distributed algorithms for two important large-scale analyses in MAGNOME's pipeline: systematic identification of gene fusion and fission events in eukaryote genomes (following [7]), and large-scale consensus clustering for protein families (following [9]). For fusions and fissions, she defined a new MapReduce algorithm that avoids graph-based analysis (which is notoriously slow in MapReduce), to achieve both significant speed ups and excellent scaling to much larger data sets. For protein family clustering, she defined a novel iterative sampling strategy that combines parallel clustering of submatrices of pairwise relations, to succesively approximate the result of a complete clustering, without the need to store the entire matrix of relations in memory.

6.4. Inferring metabolic models

Participants: David James Sherman [correspondant], Pascal Durrens, Razanne Issa, Anna Zhukova.

In collaboration with Prof Jean-Marc Nicaud's lab at the INRA Grignon, we developed the first functional genome-scale metabolic model of an oleaginous yeast. Most work in producing genome-scale metabolic models has focused on model organisms, in part due to the cost of obtaining well-annotated genome sequences and sufficiently complete experimental data for refining and verifying the models. However, for many fungal genomes of biotechnological interest, the combination of large-scale sequencing projects and in-depth experimental studies has made it feasible to undertake metabolic network reconstruction for a wider range of organisms.

An excellent representative of this new class of organisms is *Yarrowia lipolytica*, an oleaginous yeast studied experimentally for its role as a food contaminant and its use in bioremediation and cell factory applications. As one of the hemiascomycetous yeasts completely sequenced in the Génolevures program it enjoys a high quality manual annotation by a network of experts. It is also an ideal subject for studying the role of species-specific expansion of paralogous families, a considerable challenge for eukaryotes in genome-scale metabolic construction. To these ends, we undertook a complete reconstruction of the *Y. lipolytica* metabolic network.

Methods: A draft model was extrapolated from the *S. cerevisiae* model iIN800, using *in silico* methods including enzyme conservation predicted using Génolevures and reaction mapping maintaining compartments. This draft was curated by a group of experts in *Y. lipolytica* metabolism, and iteratively improved and validated through comparison with experimental data by flux balance analysis. Gap filling, species-specific reactions, and the addition of compartments with the corresponding transport reactions were among the improvements that most affected accuracy. These methods, initially implemented in an *ad hoc* way in the *Pathtastic* software tool, have been redefined and formalized by Razanne Issa using a novel logical framework.

Results: We produced an accurate functional model for *Y. lipolytica*, MODEL1111190000 in Biomodels.net, that has been qualitatively validated against gene knockouts. This model has been enriched by Anna Zhukova with ontology terms from ChEBI and GO.

6.5. Summarized visualization of metabolic models

Participants: David James Sherman [correspondant], Anna Zhukova.

In collaboration with Romain Bourqui and Antoine Lambert of the LaBRI, we defined new strategies for exploring whole genome metabolic models. There is an inherent tension between detail and understandability in these large networks: on the one hand, detailed description of individual reactions is needed for accurate simulation, but on the other hand, high-level views of reactions are needed for describing partways in human terms. We are defining knowledge-based simplification rules, that permit the user to factor similar reactions into one "generic" reaction in order to visualize a whole pathway or compartment, while maintaining the underlying model so that the user can later "drill down" to the specific reactions if need be. New layout rules implemented in the Tulip platform are used to draw the resulting networks in a familiar way.

In collaboration with Bruno Pinaud of the LaBRI, rule-based rewriting of metabolic models was used to define these simplifications using his PORGY software tool.

6.6. Hierarchical modeling with BioRica

Participants: David James Sherman [correspondant], Rodrigo Assar Cuevas, Nicolás Loira.

A recurring challenge for *in silico* modeling of cell behavior is that experimentally validated models are so focused in scope that it is difficult to repurpose them. Hierarchical modeling is one way of combining specific models into networks. Effective use of hierarchical models requires both formal definition of the semantics of such composition, and efficient simulation tools for exploring the large space of complex behaviors.

BioRica is a high-level hierarchical modeling framework for models combining continuous and discrete components. By providing a reliable and functional software tool backed by a rigorous semantics, we hope to advance real adoption of hierarchical modeling by the systems biology community. By providing an understandable and mathematically rigorous semantics, this will make is easier for practicing scientists to build practical and functional models of the systems they are studying, and concentrate their efforts on the system rather than on the tool.

Building on previous work that formalized strategies for integrating discrete control with continuous models, Rodrigo Assar defined a new framework for BioRica models using Kaufman's Quantized State Systems (in prep.).

MASAIE Project-Team

5. New Results

5.1. Robustness and $\mathcal{R}_0$

We have obtained new results about the relation between Robustness and the basic reproduction number $\mathcal{R}_0$. It is now well admitted that the basic reproduction ratio $\mathcal{R}_0$ is a key concept in mathematical epidemiology and the literature devoted to this concept is now quite important, see [20], [40], [19], [22], [23], [24], [26], [28], [30], [34] and references therein.
This number is a threshold parameter for bifurcation of an epidemic system : for a general compartmental disease transmission model, if $\mathcal{R}_0 < 1$, the disease free equilibrium (DFE) is locally asymptotically stable; whereas, if $\mathcal{R}_0 > 1$, the DFE is unstable.
It is said in some papers that $\mathcal{R}_0$ is a measure to gauge the amount of uniform effort needed to eliminate infection from a population [22], [24], [25], [31], [30].
The concept of robustness, coming from control theory, is associated to uncertainty. Usually the parameters of a system are known within a certain margin. A question is, how some properties, e.g. stability, can be ascertained with uncertainty on the parameters. In control theory “stability margin” is an important concept. Another way to formulate this problem is to analyze the effect of perturbations, unstructured or structured. This problem is also related to the so-called pseudo-spectrum [36], [37], [35].
We found that the basic reproduction number of an epidemic system is not an accurate gauge of the distance from the Jacobian J of this system, computed at the disease free equilibrium, to the set of stable matrices (if J in unstable), respectively to the set of unstable matrices (if J is stable). The same conclusion arises for another indicator, introduced by Heestebeck et al. [24], [31], [30], the type-reproduction number.

5.2. Wolbachia and Dengue

Wolbachia is a genus of bacteria which infects arthropod species, including a high proportion of insects. It is one of the world’s most common parasitic microbes and is possibly the most common reproductive parasite in the biosphere. *Wolbachia* is a maternally inherited endosymbiont of a large number of insects and other arthropods that induces various effects on host reproductive biology. Estimated to infect more than 60% of all insect species *Wolbachia* species are present in mature eggs, but not mature sperm. Only infected females pass the infection on to their offspring. Another consequence of infection is cytoplasmic incompatibility, i.e., the inability of *Wolbachia*-infected males to successfully reproduce with uninfected females.
The successful introduction of a life-shortening strain of *Wolbachia* into the dengue vector *Aedes aegypti* that halves adult lifespan has recently been reported.
Mosquitoes carrying this *Wolbachia* strain show around a 50% reduction in adult female lifespan compared to uninfected mosquitoes. It has been reported that wMel and wMelPop-CLA strains block transmission of dengue serotype 2 (DENV-2) in *Aedes aegypti*, forming the basis of a practical approach to dengue suppression. Infection by *Wolbachia* has a triple effect : reduction of recruitment, increasing of mortality for the mosquitoes and reduction of dengue transmission.
With our colleague of Brazil (see International cooperation) we built and study different models for the introduction of *Wolbachia* in a population of *Aedes aegypti*. These models are epidemiological models with vertical transmission only, which is quite new. We found that bistabilty does exist : three equilibria are present. We show that the coexistence equilibrium is unstable. We show that the equilibrium without infection and the equilibrium with the whole population infected are asymptotically stable. Numerical experimentation shows that the basin of the second equilibrium is appreciable. This indicates that introduction of *Wolbachia* is feasible. The connection of theses models with transmission models of dengue is under investigation by the French-Brazilian team.

5.3. Bilharzia

Schistosomiasis or bilharzia is a water-borne parasitic disease that affects 200 million people and poses a treat to 600 million in more than 76 countries [39]. It is caused by blood-dwelling fluke worms of the genus *Schistosoma*. The transmission cycle requires contamination of surface water by excreta, specific freshwater snails as intermediate hosts, and human water contact [21]. Schistosome are transmitted via contact with contaminated water containing cercaria the infective stage of the parasite [39], [32].

In connection with EPLS, a research NGO based in Saint-Louis (Senegal), and Pasteur Institute of Lille, we investigate a spatially deterministic metapopulation model in which infectious agents persist within a network of connected environments. This model accounts for human population age and behavior structure. We completely analyses the asymptotic behavior of this model. We give a formula for computing the basic reproduction ratio $\mathcal{R}_0$. If $\mathcal{R}_0 \leq 1$ we prove that the disease free equilibrium is globally asymptotically stable. If $\mathcal{R}_0 > 1$, with an hypothesis on connectedness, we prove that there exists a unique positive endemic equilibrium, which is globally asymptotically stable.

The validation of this model, using data of EPLS, is under investigation and is the subject of a Phd thesis. The defense will occur at the beginning of 2013. We explore the identification of key parameters using different kind of observers.

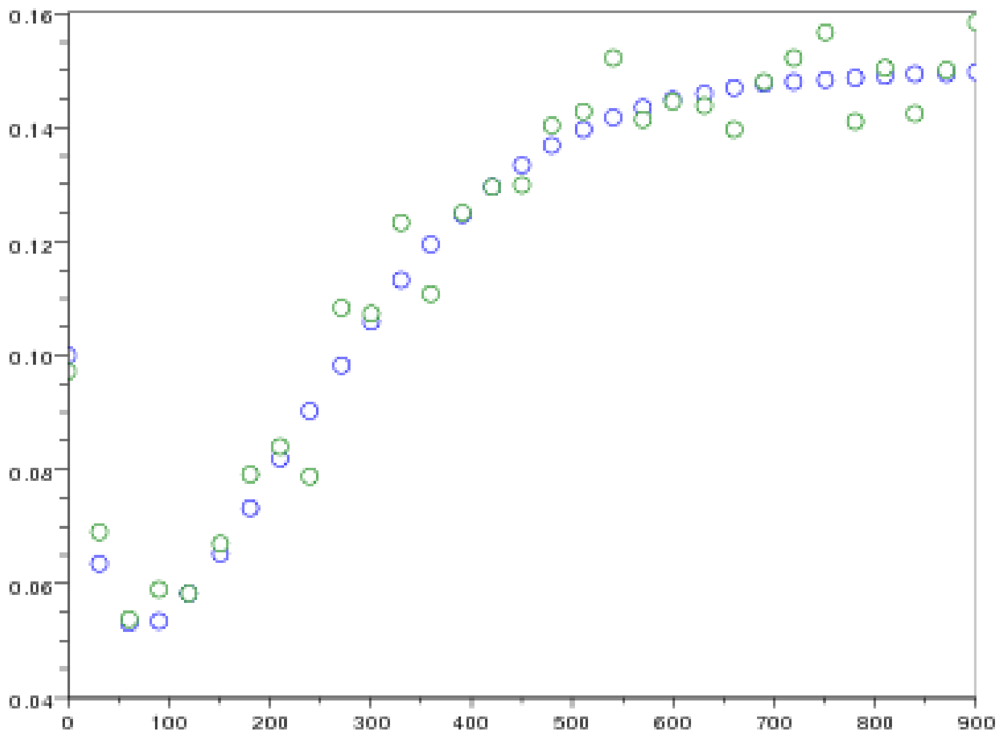


Figure 1. Noisy and discrete measure of host prevalence

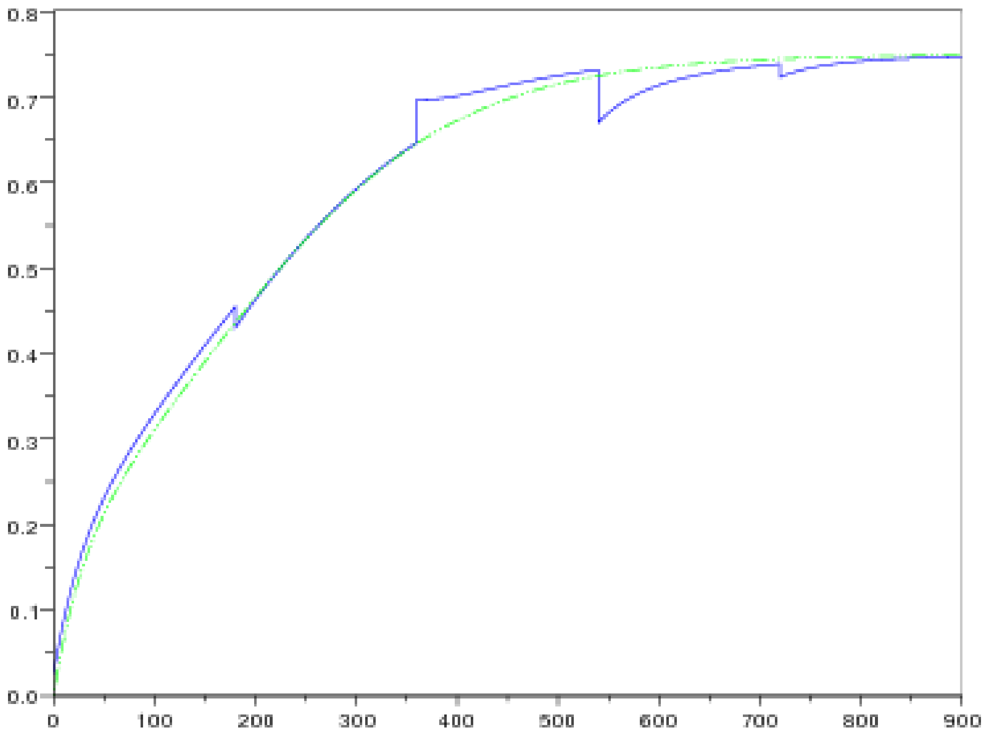


Figure 2. Reconstruction of the snail prevalence from preceding data

MNEMOSYNE Team

6. New Results

6.1. Introduction

2012 is the year of birth of the Mnemosyne team; it is also a year of transition, since most of its members were previously in the Cortex project-team in Nancy. Accordingly, this year was partly devoted to ongoing projects in the Cortex team and to the initiation of the first activities in Mnemosyne. Apart from the results listed below, corresponding to the extension of current projects from the Cortex team to our new topics, we have also begun to study models of episodic and semantic memory at the interface between the hippocampus and the cortex and interaction between models of the basal ganglia and the prefrontal cortex.

6.2. Systemic view of visuomotor transformations

Visuomotricity was an important topic in our activities in the Cortex team. We are pursuing these activities in ongoing projects (particularly the ANR Keops project, cf. § 7.1), and extend them according to a systemic view, integrating new information flows from the external world and from other neuronal structures:

- We consider the role of non-standard cells in the retina [11], reported as responsible for fast event-detection in the visual flow.
- We have initiated a modeling study of the retina-thalamus-cortex information flow and particularly of its non-specific pathway [9] that can account for attentional mechanisms.

MODEMIC Project-Team

6. New Results

6.1. Theoretical results

6.1.1. Models resource/consumer

The team maintains a significant activity about the theory of the chemostat model, proposing and studying extensions of the classical models.

6.1.1.1. Theory of competition and coexistence

Participants: Jérôme Harmand, Claude Lobry, Tewfik Sari.

In the papers [41], [50] we consider deterministic models of competition. We study the persistence of species. In [25] we study a syntrophic relation between microbial species. In [26], we give a global asymptotic stability result for a mathematical model of competition between several species in a chemostat, by using a new Lyapunov function. The model includes both monotone and non-monotone response functions, distinct removal rates for the species and variable yields, depending on the concentration of substrate.

6.1.1.2. Study of interconnected chemostats

Participants: Jérôme Harmand, Alain Rapaport.

We have shown how a particular spatial structure with a buffer globally stabilizes the chemostat dynamics with non-monotonic response function, while this is not possible with single, serial or parallel chemostats of the same total volume and input flow. We give a characterization of the set of such configurations that enjoy this property, as well as the configuration that ensures the best nutrient conversion. Furthermore, we characterize the minimal buffer volume to add to a single chemostat for obtaining the global stability. These results are illustrated with the Haldane function that models inhibition in micro-organisms growth [67].

In industrial applications, the attraction of the wash-out equilibrium is undesired because it presents a risk that may ruin the culture in case of disturbance, temporarily pump breakdown or presence of toxic material that could drive the state in the attracting basin of the wash-out equilibrium. This approach has led to a patent deposit by INRA [59] during the PhD of H. Haidar, a former PhD student of the team [80].

6.1.1.3. Aggregation models in the chemostat

Participants: Radhouene Fekih-Salem, Jérôme Harmand, Claude Lobry, Alain Rapaport, Tewfik Sari.

We have studied a model of the chemostat where the species are present in two forms, isolated and aggregated individuals, such as attached bacteria in biofilm or bacteria in flocks. We show that our general model contains a lot of models that were previously considered in the literature. Assuming that flocculation and deflocculation dynamics are fast compared to the growth of the species, we construct a reduced chemostat-like model in which both the growth functions and the apparent dilution rate depend on the density of the species. We also show that such a model involving monotonic growth rates may exhibit bi-stability, while it may occur in the classical chemostat model, but when the growth rate is non monotonic [21], [54]. This work is part of the PhD of R. Fekih-Salem co-supervised by A. Rapaport and T. Sari.

This research subject has been mainly motivated by the DISCO project (see Section 7.3).

6.1.1.4. Overyielding in continuous bioprocesses

Participants: Denis Dochain, Alain Rapaport.

We have shown that for certain configurations of two chemostats fed in parallel, the presence of two different species in each tank can improve the yield of the whole process, compared to the same configuration having the same species in each volume. This leads to a (so-called) “transgressive over-yielding” due to spatialization [35].

This work has been achieved during the stay of Prof. P. de Leenheer (Univ. Florida).

6.1.2. Measuring taxonomic diversity of microbial communities

Participant: Bart Haegeman.

Diversity is considered to be a main determinant of the behavior of microbial communities. However, measuring microbial diversity is challenging. Although metagenomic techniques allow us to sample microbial communities at unprecedented depths, the disparity between community (e.g., 10^{15} organisms) and sample (e.g., 10^5 organisms) remains large. We have studied what the diversity observed in a sample tells us about the real diversity of the community.

For a given empirical sample the aim is to construct the community from which this sample was taken. It turns out that a large set of community structures are consistent with the sample data. Some diversity metrics vary widely over this set of consistent communities, and are therefore difficult to infer from the sample data. Other diversity metrics are approximately constant over the set of consistent communities, and are therefore much easier to infer from the sample data.

The analysis of the set of consistent communities has yielded the following insights. First, it is impossible to robustly estimate the number of species from sample data. This is easy to understand. Microbial communities typically contain a large number of rare species, and these rare species are unlikely to be present in the sample. Hence, sample data are lacking crucial information to estimate species richness. Second, other diversity metrics, in particular Shannon and Simpson diversity, can be robustly estimated from sample data. We have constructed lower and upper estimates for a general class of diversity metrics, and we have shown that the difference between the extremal estimators, that is, the estimation uncertainty, is small for Shannon and Simpson diversity.

6.1.3. A theory of genetic diversity within bacterial species

Participant: Bart Haegeman.

With the wide availability of DNA sequencing, microbiologists are now able to rapidly sequence entire bacterial genomes. Comparison of these genomes has revealed a large genetic diversity within bacterial species. For example, one genome of the bacteria *E. coli* has about 4000 different genes, but a set of 10 genomes of *E. coli* has typically over 10000 different genes. Some of these genes are shared by all or almost all of the genomes, but many other genes are only present in one or a few of the genomes. This observation has important implications for the definition of bacterial species and for the description of the functional characteristics of bacteria.

In [23] we propose a theory for the frequency distribution of genes within a set of genomes. The model describes the genetic diversity as a balance between two forces. Demographic processes such as division and death tend to reduce the genetic diversity; horizontal gene transfer from other species can increase the genetic diversity. Our model predicts that the gene frequency distribution is U-shaped, meaning that there are a large number of genes present in only a few genomes, a small number of genes present in about half of the genomes, and a large number of genes present in almost all genomes. This prediction is consistent with the gene frequency distributions of six bacterial species we have analyzed (about 100 sequenced genomes in total). Importantly, the model does not assume any functional difference between the genes, that is, genes are considered to be selectively neutral. By showing that empirical gene frequency distributions can be reproduced by a neutral genome model, we contend that the frequency of a gene should not be interpreted as an indication of its essentiality or importance.

6.1.4. Individual-based modelling for bacterial ecosystems

Participants: Fabien Campillo, Chloé Deygout, Coralie Fritsch, Jérôme Harmand, Marc Joannides, Claude Lobry.

In terms of computational modelling of ecosystems, individual-based models (IBMs) are an interesting path to explore. We can outline two types of IBMs. On the one hand "detailed IBM" attempt to integrate in an ad-hoc way all the knowledge available about an ecosystem. On the other hand, "simplified IBM" are limited to one or several mechanisms to simplify the analysis. The former may be more realistic but are often difficult to analyze. Although the latter are too simplistic in realistic situations they lend themselves to the analysis and numerical analysis. We focus on the latter.

The IBMs offer an interdisciplinary language between biologists, biotechnologists, mathematicians, and computer scientists, to develop models in the form of relatively simple rules. In the case of simplified IBMs it is possible to translate these rules in the form of a branching Markov process with values in a space of measures. Using scaling methods, the IBMs can be approximated by integro-differential equations; using model simplification methods IBMs can be reduced to stochastic or ordinary differential equations. The mathematical interpretation of the IBMs and their analysis is relatively recent and still very few studies exist [78]. The numerical analysis of these models is yet to be built. Under certain conditions, IBMs themselves can be simulated through adapted Monte Carlo procedures.

The MODEMIC project-team develops many studies in the field of IBMs. The first is part of the ANR MODECOL on the modelling of clonal plant growth (see Section 7.4); the second is part of the ANR DISCO on modelling of biofilms (see Section 7.3); the third is also part of the ANR DISCO is dedicated to the modelling of biofilms in plug-flow reactors (see Section 6.2.2); the last one is the ongoing thesis of Coralie Fritsch at the École Doctorale I2E of the University of Montpellier 2, under the supervision of Fabien Campillo and Jérôme Harmand. The thesis aims at developing and analyzing "simple" individual-based microbial ecosystems models.

In all cases, we aim at developing the Monte Carlo simulation of the IBM as well as analyzing their links with integro-differential models. We also seek to make connections with non-IBM models proposed in Section 6.1.5 .

6.1.5. Stochastic/discrete and stochastic/continuous modelling for biotechnology and population dynamics

Participants: Fabien Campillo, Marc Joannides, Claude Lobry.

In [14], we consider a stochastic model of the two-dimensional chemostat as a diffusion process for the concentration of substrate and the concentration of biomass. The model allows for the washout phenomenon: the disappearance of the biomass inside the chemostat. We establish the Fokker-Planck equation associated with this diffusion process, in particular we describe the boundary conditions that modelize the washout. We propose an adapted finite difference scheme for the approximation of the solution of the Fokker-Planck equation.

In [15], we consider a hybrid version of the classical predator-prey differential equation model. The proposed model is hybrid: continuous/discrete and deterministic/stochastic. It contains a parameter ω which represents the number of individuals for one unit of prey – if x denotes the quantity of prey in the differential equation model $x = 1$ means that there are ω individuals in the discrete model – is derived from the classical birth and death process. It is shown by the mean of simulations and explained by a mathematical analysis based on results in singular perturbation theory (the so called theory of Canards) that qualitative properties of the model like persistence or extinction are dramatically sensitive to ω. This means that we must be very cautious when we use continuous variables in place of jump processes in dynamic population.

6.1.6. Optimal control of continuous bioprocesses

Participants: Walid Bouhafs, Amel Ghouali, Jérôme Harmand, Alain Rapaport.

In continuous bioprocesses, a usual objective is to stabilize the output of the bioreactors about a desired steady state (in wastewater industry, this value is typically chosen under the norm of authorized discharge). It happens more and more frequently that transient trajectories are expected also to maximize a product of interest.

We have begun to study the maximization of the gaseous production of methane in anaerobic processes over a given period of time on specific problems. For the moment we have proven that the optimal trajectory consists in approaching a unique singular arc as fast as possible when only one limiting substrate has to be converted, but the problem is still open when involving several substrates [39]. Another problem arises for alternating aerobic-anoxic systems. Revisiting the results obtained several years ago within the framework of Djalel Mazouni's thesis, we aims at proposing optimal time control policies for sequencing batch reactors in which simultaneous nitrification and denitrification take place. The solution for the original problem is a difficult task but we have already proposed a solution for a slightly modified model [30]. These last results have been obtained within the framework of the PhD thesis of W. Bouhafs.

Reference points in batch processes can be mimicked by a series of continuously stirred bioreactors in series at steady state (see applications 6.2.4 and 7.1). We study the minimal time problem to drive the nutrients concentrations of a cascade of chemostats. The control variable is the dilution rates of each tank, under the constraint that each dilution rate is bounded by the one of the previous tank, that makes the system not locally controllable. For the particular case of two tanks with total mass at steady state, the planar feedback synthesis has been found but the problem is still under investigation for the general case.

One important issue in bioprocesses is to find optimal feedback control laws in order to steer a system describing a perfectly-mixed bioreactor to a given target value in a minimal amount of time. Finding adequate feeding strategies can significantly improve the performance of the system. A typical target (for fed-batch bioreactors) is to consider the substrate concentration to be lower than a given reference value at the end of the process. Other criterium can be studied such as the maximization of the production of biomass in a given time period. Singular strategies (in reference to the theory of singular arcs in optimal control theory) can be defined in this context and are natural due to the constraints on the system. One objective of our research is to characterize optimal feedback control laws using mathematical tools from optimal control laws, and also to develop numerical methods that can handle problems with a large number of parameters.

6.1.7. Modelling the functioning of soil ecosystems

Participants: Céline Casenave, Jérôme Harmand, Alain Rapaport.

In ecology, one of the important challenges is the understanding of the biodiversity observed in the natural ecosystems. Several models have been proposed to explain this biodiversity, and in particular the coexistence of different species. In these models, it is often assumed that, when they die, the micro-organisms are directly converted in an assimilable resource, that is a resource that alive organisms can consume. However, we know that it is not the case in reality. Indeed, the organic matter stemmed from the dead organisms has to be transformed in assimilable resource before it can be consumed. This transformation is performed by some micro-organisms which have this special ability.

We have proposed a new model of soil ecosystems, of chemostat type. This model is rather simple, but also original because it takes into account several mechanisms:

- the growth, the mortality and the respiration,
- the ability of some organisms to transform the non assimilable resources in assimilable ones,
- the advantage that an organism can gain from this ability of transformation.

For the moment, we have considered the case where only one or two types of organisms are present. The model is finally composed of 3 (or 4) nonlinear ordinary differential equations: one per type of organisms, one for the assimilable resource and one for the non assimilable one. The study of the equilibrium points has first shown the possibility of coexistence, at equilibrium, of the two organisms. Then, in numerical simulations, we have observed the possible existence of limit cycles, which can also explain the coexistence of organisms observed in the nature.

This problem is still under study; a working group (in particular with researchers of the UMR Eco & Sols, conducted by B. Jaillard) meet regularly to discuss about the problems of modelling in ecology.

We have also investigated the *sampling effect* that occurs when micro-biologists select randomly species in a natural ecosystem for studying reconstituted ecosystems in a controlled environment. We have proposed a very simple probabilistic model, that shows that observing average increases or decreases on the performances of these reconstitute ecosystems when modifying the size of the sampling, allows to infer kinds and proportions of the interactions among species present in the original ecosystem [65]. This research is conducted with the UMR Eco & Sols, Montpellier, (B. Jaillard) and the UMR BIOEMCO, Grignon (N. Nunan).

6.1.8. Equivalence between models of fractured porous media

Participants: Céline Casenave, Jérôme Harmand, Alain Rapaport, Alejandro Rojas-Palma.

In geosciences, models of fractured porous media are often described as a *mobile* zone driven by advection, and one or several *immobile zones* directly or indirectly connected to the mobile zone by diffusion terms. We believe that these models are also relevant to describe flows in soil or in porous media such as biofilms. They are very close from the spatial representations used in Section 6.2.6 . We have shown, using a transfer function approach, that two schemes often used in the literature: the MINC (Multiple INteractive Continua) where diffusive compartments are connected in series, and the MRMT (Multiple Rate Mass Transfer) where diffusive compartments are connected in star around the mobile zone, are equivalent input-output representations [56], and providing formulas (up to three compartments) to pass from one representation to another. This result means that one can simply choose the most convenient approach when dealing with control or optimization without any loss of generality. We are currently working on the general case of n compartments with n larger than three, and looking for equivalent classes of configurations that could be half way between MINC and MRMT and fit better the spatial representations of real world.

This work is performed with the UMR GéoSciences Rennes (J.R. de Dreuzy), and has led to the internship of a Chilean MsC student (A. Rojas-Palma).

6.1.9. Non-linear filtering for the chemostat

Participants: Boumediene Benyahia, Amine Boutoub, Fabien Campillo, Jérôme Harmand.

Monte Carlo-based inference methods like particle filtering are bound to develop in the context of biotechnology. In contrast with the classical observer approach, inference through Monte Carlo methods can handle measurements in discrete time in low frequency and with low signal-to- noise ratio. Based on the stochastic modeling of the chemostat, these approaches may also be used for model selection and hypothesis testing.

In a preliminary work [28] we consider the bootstrap particle filter applied to a 2-dimensional chemostat model. The internship of Amine Boutoub dedicated to the study of particle filtering for more realistic chemostat models has recently started.

6.1.10. Functional identification of growth functions in the chemostat

Participant: Alain Rapaport.

We have proposed an adaptive control law that allows one to identify unstable steady states of the open-loop system in the single-species chemostat model without the knowledge of the growth function. We have then shown how to use a continuation technique to reconstruct the whole graph of the growth function, providing a new method for identifying non-monotonic growths [42], [27]. Two variants, in continuous and discrete time, have been studied. An analysis of the case of two species in competition has shown the ability of the method to detect a mixed culture for which dominance depends on the dilution rate, due to a property of stability loss in slow-fast dynamics. This method leaves open future extensions for extremum seeking problems.

This work has been conducted in cooperation with Universities of Exeter (J. Sieber) and Plymouth (S. Rodrigues), and the EPI SISYPHE (M. Desroches).

6.1.11. Diffusive representation of integro-differential models

Participant: Céline Casenave.

This work is done in collaboration with Emmanuel Montseny (LAAS/CNRS), Gérard Montseny (LAAS/CNRS), and Christophe Prieur (LIAFA/CNRS).

In lots of dynamic systems of Physics or others scientific fields such as Biology (Volterra models), dynamic integral operators, often of convolution type, are involved. Problems relating to integro-differential models are often difficult to solve, especially because these models are not time-local. In this context, the methodology called "diffusive representation" presents some interests: an integral operator is represented by its gamma-symbol, directly deduced from its transfer function. It can be formulated by means of a state realization whose dimension is numerically reasonable whatever the size of the system may be. In addition to this interesting practical side, the diffusive representation offers a unified mathematical framework, well adapted to analysis of integral convolution operators.

Several dynamic problems can be tackled in an original and quite simple way by using the diffusive representation. In fact, all the operatorial problems of modeling, simulation, control, model identification, model reduction, etc. can be formulated in such a way that the object of the problem is the gamma-symbol of the operator solution.

Several problems are under study:

- the identification of integro-differential models [66],
- the controllability of some SISO Volterra models [63],
- the simulation and the analysis of a model of porous media[64].

These works follow up on the ones developed during the PhD thesis of Céline Casenave, which deals with the problem of the operator inversion for the application to non local dynamic problems.

6.2. Applications

6.2.1. Modelling and control of Anaerobic Digestion processes

Participants: Boumediene Benyahia, Amine Charfi, Radhouene Fekih-Salem, Jérôme Harmand, Guilherme Pimentel, Tewfik Sari.

We consider the AM2 or AMOCO model developed in [72] and extend both the model in itself and its analysis to the following cases:

- Depending on the AM2 model parameters, the steady states were analytically characterized and their stability were analyzed [12]. Following this study, it was shown that the overloading tolerance, a parameter proposed in [81] to on-line monitoring anaerobic processes, may be not adapted under certain operating conditions and even lead to bad operating decisions.
- Within the framework of the PhD theses of Amine Charfi and Boumediene Benyahia, we have included the fouling dynamics of membranes into the AM2 and we have analyzed the resulting model (called the AM2b) [16], [29].
- We actually work towards two directions: (i) we are extending these results in including into the AM2 an additional process, *i. e.* the hydrolysis step in order to study bioprocesses treating solid waste (the resulting model being called the AM3) [36], [37]; (ii) we try to find links between complex models such as the ADM1 model and simple models such as the AM2b or the AM3 [40].

Apart from this work on the modelling of anaerobic digesters and membrane bioreactors, we have developed a number of specific simple models for control design accounting for the coupling of such processes with membrane modules in the chemostat (PhD thesis of G. Pimentel). This work aims at contributing to the efficient treatment of wastewaters produced in fish production farms. The work of G. Pimentel aims at studying the coupling of simple fouling models with the model of the chemostat in order to propose new simple models for control design.

6.2.2. Hybrid modelling of biofilms in plug-flow reactors

Participants: Fabien Campillo, Chloé Deygout, Annick Lesne, Alain Rapaport.

We have proposed a multi-scaled modelling that combines three scales: a microscopic one for the individual bacteria, a mesoscopic or "coarse-grained" one that homogenises at an intermediate scale the quantities relevant to the attachment/detachment process, and a macroscopic one in terms of substrate concentration.

Such a “hybrid” approach allows for modelling and understanding in plug-flow reactors the interplay between

- the formation of the biofilm at a microscopic scale, that starts from a small number of bacteria (thus a stochastic individual based description),
- the limitation of the biofilm, due the carrying capacity of the wall attachment, at a mesoscopic scale,
- the consumption of nutrient along the flow at a macroscopic level, as a solution of a coupled transport-reaction partial differential equation.

The numerical computation of such a model requires a software architecture that allows the simultaneous simulation of stochastic events at the bacteria scale and the continuous evolution (in space and time) of the substrate density.

This work has been conducted within the DISCO project (see Section 7.3) and the postdoctoral stay of C. Deygout hired by the project, in close collaboration with A. Lesne (LPTMC, Univ. Paris VI). A first paper on the simulation model has been published [17].

Within the DISCO project, experiments on real tubular plug-flow reactors have been simultaneously driven at IRSTEA Antony with the perspective of comparison with numerical simulations.

The multi-species case with different bacteria specialized in different environments (poor or rich in nutrient) is a work in progress.

6.2.3. Individual-based models for the bacterial degradation of the cellulose

Participants: Fabien Campillo, Chloé Deygout.

We propose an individual-based model for the degradation of one cellulose bead (dozens of micrometers in diameter) by cellulolytic bacteria. This model accounts for biofilm formation with minimal hypotheses: soluble substrate diffusion combined with bacterial chemotaxis-like movement in the liquid phase, lack of bacterial movement in the solid phase. The IBM results are qualitatively different from the main macroscopic degradation models previously used for cellulose degradation. It suggests that random and discrete processes could significantly impact the cellulose degradation dynamics by their effect on the spatial structuration of the colonized cellulose particles [44].

6.2.4. Modelling and control of cascade biosystems to mimic batch wine making processes

Participants: Térence Bayen, Céline Casenave, Jérôme Harmand, Alain Rapaport, Matthieu Sebbah.

An experimental setup of four tanks connected in series has been designed by the research unit SPO (Montpellier) for studying four physiological stages of yeast as steady state. The manipulated variables are the flow rates Q_i of each tank with the constraint $Q_i \geq Q_{i-1} \geq 0$, and the objective is to reach simultaneously four set-points in the four tanks. We are studying two kinds of control strategies:

- a linearizing feedback law that drives exponentially the dynamics to the target. This is not the fastest strategy but is has good robustness properties. Nevertheless, the inputs constraint imposes to use saturation functions that provide satisfactory convergence in simulations but that is hard to prove mathematically.
- a minimal time feedback. Due to lack of local controllability imposed by the constraint on the inputs, the optimal synthesis is not smooth with the presence of “barriers”. The input constraint $Q_i \geq Q_{i-1} \geq 0$ is unusual in optimal control problems that are linear w.r.t. to the control. The optimality of candidate singular arcs is still open for this problem.

This summer, some experiments have been made to test the first feedback law on the experimental setup. The control law seems to perform work, but other experiments should be made next year with more reliable input flow pumps.

This work was conducted as a part of the European CAFE project (Computer-Aided Food processes for control Engineering) described in Section 7.1 .

6.2.5. Modelling and control of an ice cream crystallization process

Participants: Céline Casenave, Denis Dochain.

In the ice cream industry, the type of final desired product (large cartons or ice creams on a stick) determine the viscosity at which the ice cream has to be produced. The control the viscosity of the ice cream at the outlet of a continuous crystallizer is therefore an important challenge. The problem has been studied in two steps. First, we have completed and validated on experimental data the reduced order model of the system. This model has been obtained by application of the method of moments on a population balance equation describing the evolution of the crystal size distribution. Then, we have proposed a nonlinear control strategy to control of the viscosity of the ice cream with the temperature of the refrigerant fluid of the crystallizer. This control strategy is based on a linearizing control law coupled with a Smith predictor to account for the measurement delay. The control has been validated on an experimental pilot plant located at IRSTEA (Antony, France).

This work was conducted as a part of the European CAFE project (Computer-Aided Food processes for control Engineering) described in Section 7.1 .

6.2.6. Bioremediation of natural resources

Participants: Sébastien Barbier, Jérôme Harmand, Alain Rapaport, Antoine Rousseau.

The objective of this work is to provide efficient strategies for the bioremediation of natural water resources. The originality of the approach is to couple minimal time strategies that are determined on a simplified model with a faithful numerical model for the hydrodynamics. This work has been carried out in close cooperation with A. Rousseau. Based on a previous paper that deals with an implicit representation of the spatial inhomogeneity of the resource with a small number of homogeneous compartments (with a system of ODEs), we have implemented a coupled ODE-PDE system that accounts for the spatial non-homogeneity of pollution in natural resources. The main idea is to implement a Navier-Stokes model in the resource (such as a lake), with boundary conditions that correspond to the output feedback that has been determined to be optimal for the simple ODEs model of a (small) bioreactor. A first mathematical model has been introduced and numerical simulations have been performed in academic situations. During the internship of S. Barbier (co-advised by A. Rousseau and A. Rapaport) we built a reduced model that approximates the reference PDE model thanks to a set of ODEs with parameters. Numerical optimization is performed on these parameters in order to better fit the reference model. This will lead to a publication.

The study of the minimal time strategies on the system of ODEs has been mainly achieved in cooperation with Chilean researchers (P. Gajardo, Universidad Tecnica Federico Santa Maria, and H. Ramirez, Centro de Modelamiento Matemático) and a Chilean PhD student (V. Riquelme, Depto. Ingenieria Matematica, Universidad de Chile) within the associated team DYMECOS [57].

6.2.7. Modelling and simulating terrestrial plant ecological dynamics

Participant: Fabien Campillo.

This study is part of the ANR Syscomm MODECOL that is done in collaboration particularly with the University of Rennes I, the University of La Rochelle and Inria. The first semester of 2012 was the last part of the project. We propose a very original individual-based model for clonal plant dynamics in continuous time and space that focuses on the effects of the network structure of the plants on the reproductive strategy of ramets. The model is coupled with a PDE dynamics for resources. The basic structure of the IBM encompass a population of “ramets” (the individuals) connected by “stolons or rhizomes” (the network) [13], [22]. See http://www-sop.inria.fr/members/Fabien.Campillo/software/ibm-clonal/ for more details.

6.2.8. Modelling and inferring agricultural dynamics

Participants: Fabien Campillo, Angelo Raherinirina.

The International Laboratory LIRMA supports this work that is done in collaboration with the University of Fianarantsoa in Madagascar and with Dominique Hervé (IRD, Fianarantsoa, Madagascar). The aim is to study the dynamics of agricultural plots on the edge of primary forest. In [32] a simple Markov model has been successfully confronted to a first data set with the help of maximum likelihood and Bayesian approaches. On a new data set developed by IRD, the Markov hypothesis has been rejected and we proposed to use semi-Makov models: for this new dataset the law of the sojourn time on certain states will depend on the next state visited, which is incompatible with the Markov hypothesis and which is consistent with the semi-Markov hypothesis.

MOISE Project-Team

6. New Results

6.1. Mathematical Modelling of the Ocean Dynamics

6.1.1. Beyond the traditional approximation on the Coriolis force

Participant: Antoine Rousseau.

Formerly, A. Rousseau has performed some theoretical and numerical studies around the derivation of quasi-hydrostatic models. With C. Lucas, he proved that it is sometimes necessary to take into account the cosine part of the Coriolis force (which is usually neglected, leading to the so-called Traditional Approximation). They have also shown that the non-traditional terms do not raise any additional mathematical difficulty in the primitive equations: well-posedness for both weak and strong solutions.
A. Rousseau and J. McWilliams (UCLA) proposed a mathematical justification of the tilt of convective plumes in the quasi-geostrophic regime, thanks to the account of the complete Coriolis force in the so-called quasi-hydrostatic quasi-geostrophic (QHQG) model. The new model has been presented in international conferences [59] and [60].

6.1.2. Coupling Methods for Oceanic and Atmospheric Models

Participants: Eric Blayo, David Cherel, Laurent Debreu, Antoine Rousseau, Manel Tayachi.

6.1.2.1. Interface conditions for coupling ocean models

Many physical situations require coupling two models with not only different resolutions, but also different physics. Such a coupling can be studied within the framework of global-in-time Schwarz methods. However, the efficiency of these iterative algorithms is strongly dependent on interface conditions. As a first step towards coupling a regional scale primitive equations ocean model with a local Navier-Stokes model, a study on the derivation of interface conditions for 2-D $x-z$ Navier-Stokes equations has been performed in D. Cherel PhD thesis. It has been shown theoretically that several usual conditions lead to divergent algorithms, and that a convergent algorithm is obtained when using transmission conditions given by a variational calculation.
D. Cherel has implemented a Schwarz-based domain decomposition method, for which he developed optimized absorbing boundary conditions that mix the velocity and pressure variables on an Arakawa-C grid. The numerical results confirm the rate of convergence that has been obtained theoretically, thanks to a Fourier analysis of the semi-discretized problem.
A first step towards the coupling between Navier-Stokes and primitive equations has been made in 2012. Starting from the optimized boundary conditions obtained for the Navier-Stokes equations, we performed an asymptotic analysis in order to obtain boundary conditions that should supplement the hydrostatic Navier-Stokes equations. These results have been presented in national and international conferences [47], [46], a paper is in preparation. David Cherel defended his PhD on Dec. 12th, 2012.

6.1.2.2. Coupling dimensionally heterogeneous models

The coupling of different types of models is gaining more and more attention recently. This is due, in particular, to the needs of more global models encompassing different disciplines (*e.g.* multi-physics) and different approaches (*e.g.* multi-scale, nesting). Also, the possibility to assemble different modeling units inside a friendly modelling software platform is an attractive solution compared to developing more and more global complex models. More specifically one may want to couple 1D to 2D or 3D models, such as Shallow Water and Navier Stokes models: this is the framework of our partnership with EDF in the project MECSICO. In her PhD, M. Tayachi is aimed to build a theoretical and numerical framework to couple 1D, 2D and 3D models for river flows.

In [65], we propose and analyze an efficient iterative coupling method for a dimensionally heterogeneous problem. We consider the case of a 2-D Laplace equation with non symmetric boundary conditions with a corresponding 1-D Laplace equation. We first show how to obtain the 1-D model from the 2-D one by integration along one direction, by analogy with the link between shallow water equations and the Navier-Stokes system. Then we focus on the design of a Schwarz-like iterative coupling method. We discuss the choice of boundary conditions at coupling interfaces. We prove the convergence of such algorithms and give some theoretical results related to the choice of the location of the coupling interface, and to the control of the difference between a global 2-D reference solution and the 2-D coupled one. These theoretical results are illustrated numerically. The extension of this work to shallow0water equations and primitive equations has been started recently.

6.1.3. Numerical schemes for ocean modelling

Participants: Laurent Debreu, Jérémie Demange.

Reducing the traditional errors in terrain-following vertical coordinate ocean models (or sigma models) has been a focus of interest for the last two decades. The objective is to use this class of model in regional domains which include not only the continental shelf, but the slope and deep ocean as well. Two general types of error have been identified: 1) the pressure-gradient error and 2) spurious diapycnal diffusion associated with steepness of the vertical coordinate. In a recent paper [87], we have studied the problem of diapycnal mixing. The solution to this problem requires a specifically designed advection scheme. We propose and validate a new scheme, where diffusion is split from advection and is represented by a rotated biharmonic diffusion scheme with flow-dependent hyperdiffusivity satisfying the Peclet constraint.
In 2012, in collaboration with F. Lemarié at UCLA, this work has been extended in order to render the biharmonic diffusion operator scheme unconditionally stable [17]. This is particularly needed when the slopes between coordinates lines and isopycnal surfaces are important so that the rotation of the biharmonic leads to strong stability condition along the vertical coordinate where the grid size is relatively small. This work also extends more classical results on the stability of laplacian diffusion with mixed derivatives.

In his PhD, Jérémie Demange begins a work on advection-diffusion schemes for ocean models (Supervisors : L. Debreu, P. Marchesiello (IRD)). His work will focus on the link between tracers (temperature and salinity) and momentum advection and diffusion in the non hyperbolic system of equations typically used in ocean models (the so called primitive equations with hydrostatic and Boussinesq assumptions). We also investigated the use of a depth dependent barotropic mode in free surface ocean models. When most ocean models assume that this mode is vertically constant, we have shown that the use of the true barotropic mode, derived from a normal mode decomposition, allows more stability and accuracy in the representation of external gravity waves [49], [48].

Salinity at 1000 m in the Southwest Pacific ocean is shown in figure 1 . The use of traditional upwind biased schemes (middle) exhibits a strong drift in the salinity field in comparison with climatology (left). The introduction of high order diffusion rotated along geopotential surfaces prevents this drift while maintaining high resolution features (right).

6.2. Data Assimilation for Geophysical Models

6.2.1. Development of a Variational Data Assimilation System for OPA9/NEMO

Participants: Arthur Vidard, Bénédicte Lemieux-Dudon, Pierre-Antoine Bouttier.

We are heavily involved in the development of NEMOVAR (Variational assimilation for NEMO). For several years now, we built a working group (coordinated by A. Vidard) in order to bring together various NEMOVAR user-groups with diverse scientific interests (ranging from singular vector and sensitivity studies to specific issues in variational assimilation) It has led to the creation of the VODA (Variational Ocean Data Assimilation for multi scales applications) ANR project (ended in 2012). A new project, part of a larger EU-FP7 project has been submitted late 2012.

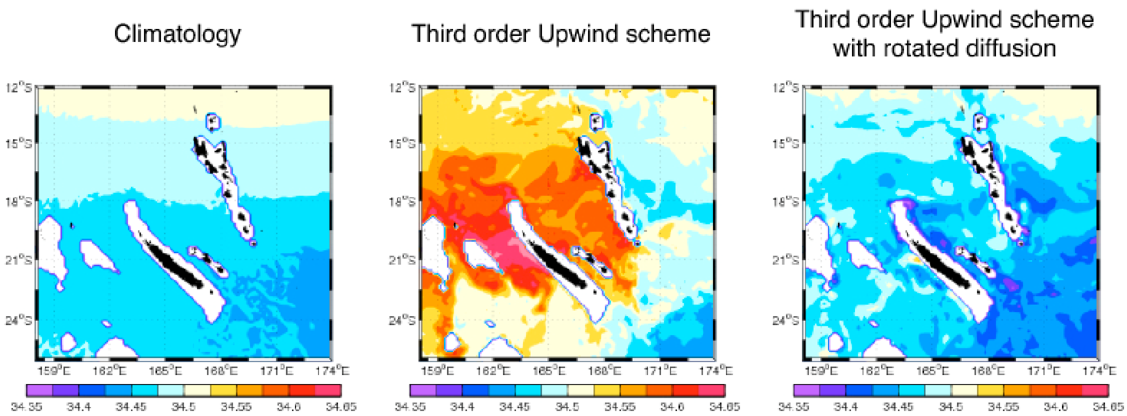


Figure 1. Salinity at 1000m in the Southwest Pacific ocean.

The project aims at delivering a common NEMOVAR platform based on NEMO platform for 3D and 4D variational assimilation. Following 2009-11 VODA activities, a fully parallel version of NEMOTAM (Tangent and Adjoint Model for NEMO) is now available for the community in the standard NEMO version. This version is based on the released 3.4.1 version of NEMO.

We are also investigating variational data assimilation methods applied to high resolution ocean numerical models. This part of the project is now well advanced and encouraging preliminary results are available on an idealized numerical configuration of an oceanic basin (see Figure 2). Several novative diagnostics have been also developed in this framework

6.2.2. Identification of pollution.

Participant: François-Xavier Le Dimet.

The problem is the next : potential sources of pollution are known but the contribution of each source to a local site is unknown. The problem is to identify the contribution of each source. This is a very common situation both at the local scale and at the synoptic scale. Thanks to second order methods we have been able to reach this goal, the theoretical part is done at FSU and application at the Institute of Mechanics of the Vietnamese Academy of Sciences. One paper has been submitted for publication [75]. At FSU M.Y. Hussaini and I. Souopgui are involved in this project

The quality of water ressources is an important problem for Vietnam. With scientists of the Institute of Mechanics (Ha Tran Thua, Hoang Van Lai, Nguyen Ba Hung) in [31] and [53] we have used the methods described in [75] for water pollution studies, in parallel Tran Thu Ha and Pham Dinh Tuan (LJK) have been working on the application of Kalman filter for this problem. Several talks have been given and papers published.

6.2.3. Variational data assimilation for large scale ice-sheet models

Participants: Bertrand Bonan, Maëlle Nodet, Catherine Ritz.

In collaboration with C. Ritz (CNRS, Laboratoire de Glaciologie et Geophysique de l'Environnement (LGGE), Grenoble), we aim to develop adjoint methods for ice cap models.

In the framework of global warming, the evolution of sea level is a major but ill-known phenomenon. It is difficult to validate the models which are used to predict the sea level elevation, because observations are heterogeneous and sparse.

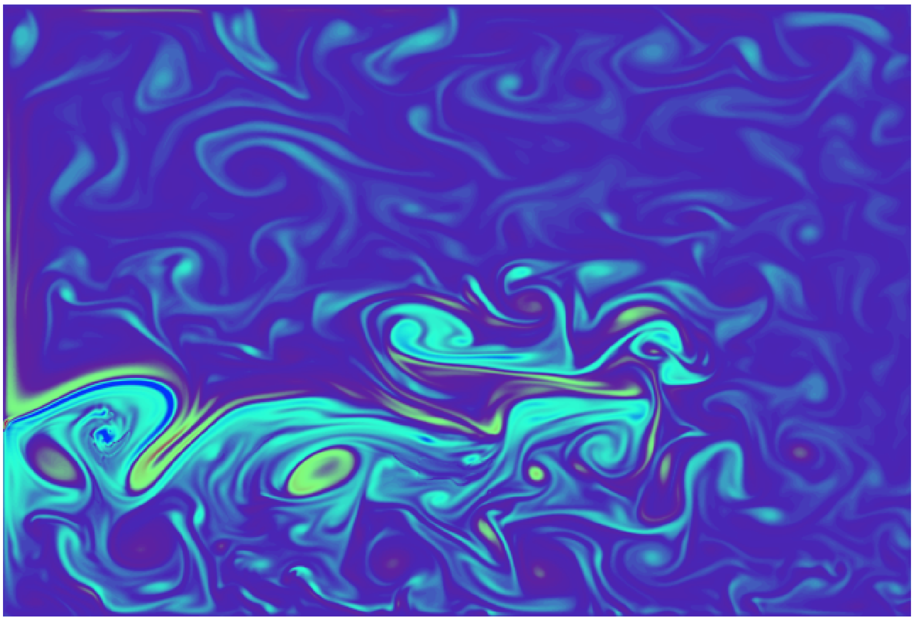


Figure 2. Surface relative vorticity of a 1/24th of a degree NEMO configuration

Data acquisition in polar glaciology is difficult and expensive. Satellite data have a good spatial coverage, but they allow only indirect observation of the interesting data. We wish to make the most of all available data and evaluate what new observations to add, where and when. Sensitivity analysis, and in particular the adjoint method, allows to identify the most influential parameters and variables and can help to design the observation network.

B. Bonan started his PhD in September 2010 on this subject. We implemented the 4D-Var algorithm for a flowline Shallow-Ice model, called Winnie, developed by C. Ritz at LGGE. In a simple configuration, we were able to generate the adjoint code by automatic differentiation. First results were encouraging and were presented at EGU [58] and Les Houches Summer School [30].

6.2.4. Ensemble Kalman filtering for large scale ice-sheet models

Participants: Bertrand Bonan, Maëlle Nodet, Catherine Ritz.

We are also interested in comparing variational methods to stochastic filtering. In the framework of B. Bonan PhD, we then implemented Ensemble Transform Kalman Filter (ETKF) on Winnie, which we would like to compare to variational assimilation methods. First results are promising and were presented at three conferences [39], [37], [38].

6.2.5. Inverse methods for full-Stokes glaciology models

Participants: Olivier Gagliardini, Maëlle Nodet, Catherine Ritz.

We are investigating the means to apply inverse modeling to another class of glaciology models, called full-Stokes model. Such a model is developed by LGGE and CSC in Finland, called Elmer/Ice. Contrary to large scale models, Elmer/Ice is based on the full Stokes equations, and no assumptions regarding aspect ratio are made, so that this model is well adapted to high resolution small scale modelling, such as glaciers (and more recently the whole Greenland ice-sheet).

In collaboration with O. Gagliardini, F. Gillet-Chaulet and C. Ritz (Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), Grenoble), we investigated a new method to solve inverse problems for a Full-Stokes model of Groenland, which consisted in solving iteratively a sequence of Neumann and Dirichlet problems within a gradient descent algorithm. We also compared this method to an approximate variational algorithm, using the fact that the full Stokes equations are almost self-adjoint. These results have been published in The Cryosphere Discussion [11] and presented at three conferences [52], [28], [57]. Figure 3 presents the reconstructed surface velocities compared to the observations, where we can see a good agreement of the main features, thus success of the assimilation process.

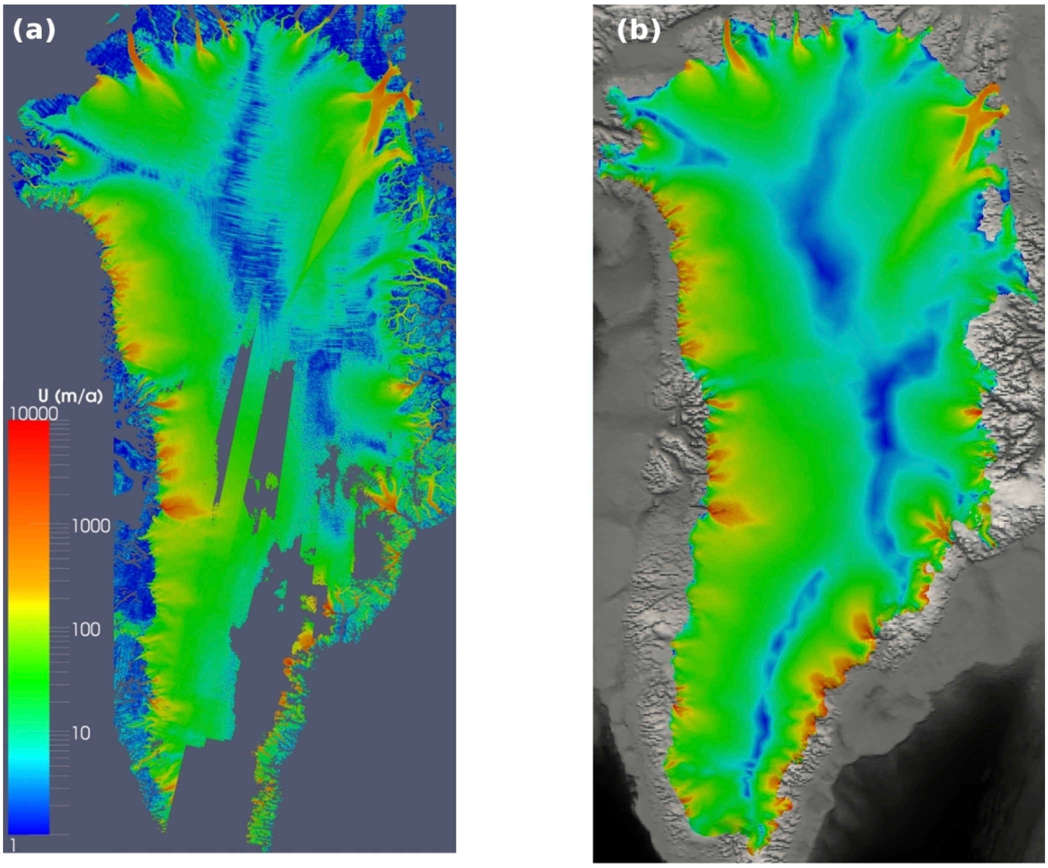


Figure 3. Surface velocities of Greenland Ice Sheet, in meters per year. On the left (a), the velocities which are observed by satellites. On the right (b), velocities obtained after assimilation.

6.2.6. Dating ice matrix and gas bubbles with DatIce

Participants: Eric Blayo, Bénédicte Lemieux-Dudon, Habib Toye Mahamadou Kele.

H. Toye Mahamadou Kele joined the MOISE team for 2 years as an Inria young engineer. A shared memory parallelization of the code and a more friendly user interface have been developed. Efforts have been made to calibrate the error covariance matrices by the mean of a posteriori diagnostics.

The MOISE team was involved in the Antarctic Ice Core Chronology 2012 (AICC2012) through a tight collaboration with the Laboratoire de Glaciologie et de Géophysique de l'Environnement (LGGE), the

Laboratoire des Sciences du Climat et de l'Environnement (LSCE), and other European laboratories. The AICC2012 project aimed at constructing an unified chronology for several Antarctic ice cores. A special issue is dedicated to AICC2012 in Climate of the Past http://www.clim-past-discuss.net/special_issue53.html. MOISE efforts on the DatIce code lead to two important articles currently reviewed in the Open Discussion process http://www.clim-past-discuss.net/papers_in_open_discussion.html: [2] and [22].

6.3. Development of New Methods for Data Assimilation

6.3.1. Variational Data Assimilation with Control of Model Error

Participants: Bénédicte Lemieux-Dudon, Arthur Vidard.

One of the main limitation of the current operational variational data assimilation techniques is that they assume the model to be perfect mainly because of computing cost issues. Numerous researches have been carried out to reduce the cost of controlling model errors by controlling the correction term only in certain privileged directions or by controlling only the systematic and time correlated part of the error.

Both the above methods consider the model errors as a forcing term in the model equations. Trémolet (2006) describes another approach where the full state vector (4D field: 3D spatial + time) is controlled. Because of computing cost one cannot obviously control the model state at each time step. Therefore, the assimilation window is split into sub-windows, and only the initial conditions of each sub-window are controlled, the junctions between each sub-window being penalized. One interesting property is that, in this case, the computation of the gradients, for the different sub-windows, are independent and therefore can be done in parallel.

This method is now implemented in a realistic Oceanic framework using OPAVAR/ NEMOVAR. An extensive documentation has been produced and we are now assessing the improvement over the previous scheme

6.3.2. Variational Data Assimilation and Control of Boundary Conditions

Participant: Eugène Kazantsev.

A variational data assimilation technique applied to the identification of the optimal discretization of interpolation operators and derivatives in nodes that are adjacent to the boundary of the domain is discussed in two contexts: a simplified case of a shallow water model and the ORCA-2 configuration of the NEMO model.

Experiments with a non-linear shallow water model in [14] show that controlling the discretization of operators near a rigid boundary can bring the model solution closer to observations both within and beyond the assimilation window. This type of control allows also to improve climatic variability of the model. These properties have been studied in two different configurations: an academic case of assimilation of artificially generated observational data in a square box configuration and assimilation of real observations in a model of the Black sea.

The sensitivity of the shallow water model in the previously described configurations has been studied in detail in [15]. It is shown in both experiments that boundary conditions near a rigid boundary influence the solution higher than the initial conditions. This fact points out the necessity to identify optimal boundary approximation during a model development.

Considering a full-physics global ocean model, we apply the 4D-Var data assimilation technique to ORCA-2 configuration of the NEMO in order to identify the optimal parametrization of boundary conditions on the lateral boundaries as well as on the bottom and on the surface of the ocean [71]. The influence of boundary conditions on the solution is analyzed as in the assimilation window and beyond the window. It is shown that optimal surface and bottom boundary conditions allow us to better represent the jet streams, such as Gulf Stream and Kuroshio. Sea Surface Height in the North Atlantic before and after control is shown in fig.4). Analyzing the reasons of the jets reinforcement, we notice that data assimilation has a major impact on parametrization of the bottom boundary conditions for u and v [55].

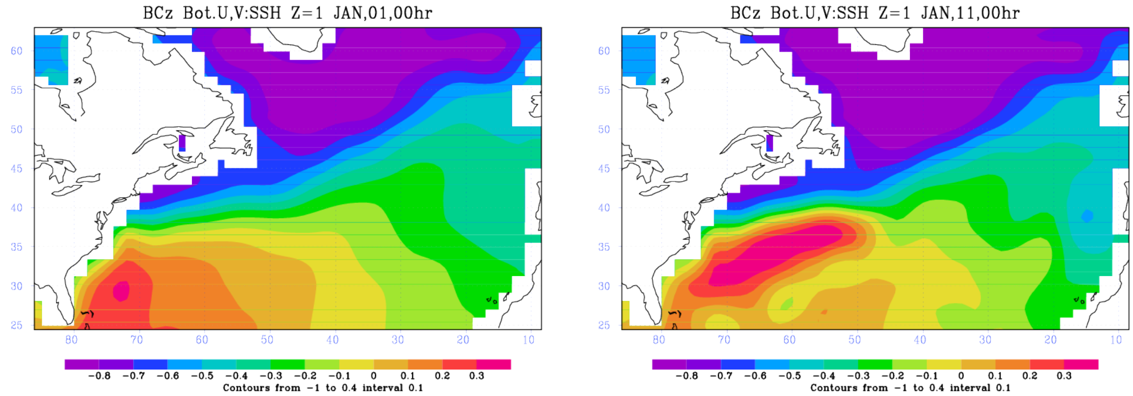


Figure 4. Sea surface elevation in the North Atlantic on the 1st (classical boundary) and on the 11th (optimal boundary) of January, 2006.

Adjoint models, necessary to variational data assimilation have been produced by the TAPENADE software, developed by the TROPICS team. This software is shown to be able to produce the adjoint code, that can be used in data assimilation after a memory usage optimization.

6.3.3. Direct assimilation of sequences of images

Participants: François-Xavier Le Dimet, Maëlle Nodet, Nicolas Papadakis, Arthur Vidard, Vincent Chabot.

At the present time the observation of Earth from space is done by more than thirty satellites. These platforms provide two kinds of observational information:

- Eulerian information as radiance measurements: the radiative properties of the earth and its fluid envelops. These data can be plugged into numerical models by solving some inverse problems.
- Lagrangian information: the movement of fronts and vortices give information on the dynamics of the fluid. Presently this information is scarcely used in meteorology by following small cumulus clouds and using them as Lagrangian tracers, but the selection of these clouds must be done by hand and the altitude of the selected clouds must be known. This is done by using the temperature of the top of the cloud.

MOISE was the leader of the ANR ADDISA project dedicated to the assimilation of images, and is a member of its current follow-up GeoFluids (along with EPI FLUMINANCE and CLIME, and LMD, IFREMER and Météo-France)

During the ADDISA project we developed Direct Image Sequences Assimilation (DISA) and proposed a new scheme for the regularization of optical flow problems [95]. Thanks to the nonlinear brightness assumption, we proposed an algorithm to estimate the motion between two images, based on the minimization of a nonlinear cost function. We proved its efficiency and robustness on simulated and experimental geophysical flows [78]. As part of the ANR project GeoFluids, we are investigating new ways to define distance between a couple of images. One idea is to compare the gradient of the images rather than the actual value of the pixels. This leads to promising results. Another idea, currently under investigation, consists in comparing mains structures within each image. This can be done using, for example, a wavelet representation of images. We are also part of TOMMI, another ANR project started mid 2011, where we are investigating the possibility to use optimal transportation based distances for images assimilation.

6.3.4. Assimilation of ocean images

Participants: Vincent Chabot, Maëlle Nodet, Nicolas Papadakis, Arthur Vidard, Alexandros Makris.

In addition with the direct assimilation approach previously described, a particular attention has been given to the study of data noise for ocean image assimilation. A journal paper is about to be submitted on this subject in Tellus A. In the context of assimilation of structures contained in satellite images, a two step registration approach using computer vision tools has been proposed within the post-doctorate of Alexandros Makris [33], [56].

6.4. Quantifying Uncertainty

6.4.1. Sensitivity analysis for West African monsoon

Participants: Anestis Antoniadis, Céline Helbert, Clémentine Prieur, Laurence Viry.

6.4.1.1. Geophysical context

The West African monsoon is the major atmospheric phenomenon which drives the rainfall regime in Western Africa. Therefore, this is the main phenomenon in water resources over the African continent from the equatorial zone to the sub-Saharian one. Obviously, it has a major impact on agricultural activities and thus on the population itself. The causes of inter-annual spatio-temporal variability of monsoon rainfall have not yet been univocally determined. Spatio-temporal changes on the see surface temperature (SST) within the Guinea Gulf and Saharian and Sub-Saharian Albedo are identified by a considerable body of evidences as major factors to explain it.

The aim of this study is to simulate the rainfall by a regional atmospheric model (RAM) and to analyze its sensitivity to the variability of these inputs parameters. Once precipitations from RAM are compared to several precipitation data sets we can observe that the RAM simulates the West African monsoon reasonably.

6.4.1.2. Statistical methodology

As mentioned in the previous paragraph, our main goal is to perform a sensitivity analysis for the West African monsoon. Each simulation of the regional atmospheric model (RAM) is time consuming, and we first have to think about a simplified model. We deal here with spatio-temporal dynamics, for which we have to develop functional efficient statistical tools. In our context indeed, both inputs (albedo, SST) and outputs (precipitations) are considered as time and space indexed stochastic processes. A first step consists in proposing a functional modeling for both precipitation and sea surface temperatures, based on a new filtering method. For each spatial grid point in the Gulf of Guinea and each year of observation, the sea surface temperature is measured during the active period on a temporal grid. A Karhunen-Loève decomposition is then performed at each location on the spatial grid [97]. The estimation of the time dependent eigenvalues at different spatial locations generates great amounts of high-dimensional data. Clustering algorithms become then crucial in reducing the dimensionality of such data.

Thanks to the functional clustering performed on the first principal component at each point, we have defined specific subregions in the Gulf of Guinea. On each subregion, we then choose a referent point for which we keep a prescribed number of principal components which define the basis functions. The sea surface temperature at any point in this subregion is modeled by the projection on this truncated basis. The spatial dependence is described by the coefficients of the projection. The same approach is used for precipitation. Hence for both precipitation and sea surface temperatures, we obtain a decomposition where the basis functions are functions depending on time and whose coefficients are spatially indexed and time independent. Then, the most straightforward way to model the dependence of precipitation on sea surface temperatures is through a multivariate response linear regression model with the output (precipitation) spatially indexed coefficients in the above decomposition and the input (SST) spatially indexed coefficients being predictors. A naive approach consists in regressing each response onto the predictors separately; however it is unlikely to produce satisfactory results, as such methods often lead to high variability and over-fitting. Indeed the dimensions of both predictors and responses are large (compared to the sample size).

We apply a novel method recently developed by [91] in integrated genomic studies which takes into account both aspects. The method uses an ℓ_1-norm penalty to control the overall sparsity of the coefficient matrix of the multivariate linear regression model. In addition, it also imposes a *group* sparse penalty. This penalty puts a constraint on the ℓ_2 norm of regression coefficients for each predictor, which thus controls the total number of predictors entering the model, and consequently facilitates the detection of important predictors. The dimensions of both predictors and responses are large (compared to the sample size). Thus in addition to assuming that only a subset of predictors enter the model, it is also reasonable to assume that a predictor may affect only some but not all responses. By the way we take into account the complex and spatio-temporal dynamics. This work has been published in [1].

6.4.1.3. Distributed Interactive Engineering Toolbox

An important point in the study described above is that the numerical storage and processing of model inputs/outputs requires considerable computation resources. They were performed in a grid computing environment with a middleware (DIET) which takes into account the scheduling of a huge number of computation requests, the data-management and gives a transparent access to a distributed and heterogeneous platform on the regional Grid CIMENT (http://ciment.ujf-grenoble.fr/).

Thus, a different DIET module was improved through this application. An automatic support of a data grid software (http://www.irods.org) through DIET and a new web interface designed for MAR was provided to physicians.
These works involve also partners from the Inria project/team GRAAL for the computational approach, and from the Laboratory of Glaciology and Geophysical Environment (LGGE) for the use and interpretation of the regional atmospheric model (RAM).

6.4.2. Tracking for mesoscale convective systems

Participants: Anestis Antoniadis, Céline Helbert, Clémentine Prieur, Laurence Viry, Roukaya Keinj.

6.4.2.1. Scientific context

In this section, we are still concerned with the monsoon phenomenon in western Africa and more generally with the impact of climate change. What we propose in this study is to focus on the analysis of rainfall system monitoring provided by satellite remote sensing. The available data are micro-wave and IR satellite data. Such data allow characterizing the behavior of the mesoscale convective systems. We wish to develop stochastic tracking models, allowing for simulating rainfall scenari with uncertainties assessment.

6.4.2.2. Stochastical approach

The chosen approach for tracking these convective systems and estimating the rainfall intensities is a stochastic one. The stochastic modeling approach is promising as it allows developping models for which confidence in the estimates and predictions can be evaluated. The stochastic model will be used for hydro-climatic applications in West Africa. The first part of the work will consist in implementing a model developed in [96] on a test set to evaluate its performances, our ability to infer the parameters, and the meaning of these parameters. Once the model well fitted on toy cases, this algorithm should be run on our data set, and compared with previous results by [89] or by [88]. The model developed by [96] is a continuous time stochastic model to multiple target tracking, which allows in addition to birth and death, splitting and merging of the targets. The location of a target is assumed to behave like a Gaussian Process when it is observable. Targets are allowed to go undetected. Then, a Markov Chain State Model decides when the births, death, splitting or merging of targets arise. The tracking estimate maximizes the conditional density of the unknown variables given the data. The problem of quantifying the confidence in the estimate is also addressed. Roukaya Keinj started working on this topic with a two years postdoctoral position in November 2011. She left the team in October 2012, and is now replaced by Alexandros Makris.

6.4.3. Sensitivity analysis for forecasting ocean models

Participants: Anestis Antoniadis, Eric Blayo, Gaëlle Chastaing, Céline Helbert, Alexandre Janon, François-Xavier Le Dimet, Simon Nanty, Maëlle Nodet, Clémentine Prieur, Jean-Yves Tissot, Federico Zertuche.

6.4.3.1. Scientific context

Forecasting ocean systems require complex models, which sometimes need to be coupled, and which make use of data assimilation. The objective of this project is, for a given output of such a system, to identify the most influential parameters, and to evaluate the effect of uncertainty in input parameters on model output. Existing stochastic tools are not well suited for high dimension problems (in particular time-dependent problems), while deterministic tools are fully applicable but only provide limited information. So the challenge is to gather expertise on one hand on numerical approximation and control of Partial Differential Equations, and on the other hand on stochastic methods for sensitivity analysis, in order to develop and design innovative stochastic solutions to study high dimension models and to propose new hybrid approaches combining the stochastic and deterministic methods.

6.4.3.2. Estimating sensitivity indices

A first task is to develop tools for estimated sensitivity indices. Among various tools a particular attention was first paid to FAST and its derivatives. In [21], the authors present a general way to correct a positive bias which occurs in all the estimators in random balance design method (RBD) and in its hybrid version, RBD-FAST. Both these techniques derive from Fourier amplitude sensitivity test (FAST) and, as a consequence, are faced with most of its inherent issues. And up to now, one of these, the well-known problem of interferences, has always been ignored in RBD. After presenting in which way interferences lead to a positive bias in the estimator of first-order sensitivity indices in RBD, the authors explain how to overcome this issue. They then extend the bias correction method to the estimation of sensitivity indices of any order in RBD-FAST. They also give an economical strategy to estimate all the first-order and second-order sensitivity indices using RBD-FAST. A more theoretical work [77] revisit FAST and RBD in light of the discrete Fourier transform (DFT) on finite subgroups of the torus and randomized orthogonal array sampling. In [77] the authors study the estimation error of both these methods. This allows to improve FAST and to derive explicit rates of convergence of its estimators by using the framework of lattice rules. A natural generalization of the classic RBD is also provided, by using randomized orthogonal arrays having any parameters, and a bias correction method for its estimators is proposed. In variance-based sensitivity analysis, another classical tool is the method of Sobol' [94] which allows to compute Sobol' indices using Monte Carlo integration. One of the main drawbacks of this approach is that the estimation of Sobol' indices requires the use of several samples. For example, in a d-dimensional space, the estimation of all the first-order Sobol' indices requires $d + 1$ samples. Some interesting combinatorial results have been introduced to weaken this defect, in particular by Saltelli [93] and more recently by Owen [90] but the quantities they estimate still require $O(d)$ samples. In a recent work [76] the authors introduce a new approach to estimate for any k all the k-th order Sobol' indices by using only two samples based on replicated latin hypercubes. They establish theoretical properties of such a method for the first-order Sobol' indices and discuss the generalization to higher-order indices. As an illustration, they propose to apply this new approach to a marine ecosystem model of the Ligurian sea (northwestern Mediterranean) in order to study the relative importance of its several parameters. The calibration process of this kind of chemical simulators is well-known to be quite intricate, and a rigorous and robust — i.e. valid without strong regularity assumptions — sensitivity analysis, as the method of Sobol' provides, could be of great help.

6.4.3.3. Intrusive sensitivity analysis, reduced models

Another point developed in the team for sensitivity analysis is model reduction. To be more precise regarding model reduction, the aim is to reduce the number of unknown variables (to be computed by the model), using a well chosen basis. Instead of discretizing the model over a huge grid (with millions of points), the state vector of the model is projected on the subspace spanned by this basis (of a far lesser dimension). The choice of the basis is of course crucial and implies the success or failure of the reduced model. Various model reduction methods offer various choices of basis functions. A well-known method is called proper orthogonal decomposition" or principal component analysis". More recent and sophisticated methods also exist and may be studied, depending on the needs raised by the theoretical study. Model reduction is a natural way to overcome difficulties due to huge computational times due to discretizations on fine grids. In [12], the authors present a reduced basis offline/online procedure for viscous Burgers initial boundary value problem, enabling efficient approximate computation of the solutions of this equation for parametrized viscosity and initial and boundary value data. This procedure comes with a fast-evaluated rigorous error bound certifying the approximation procedure. The numerical experiments in the paper show significant computational savings, as well as efficiency of the error bound. The present preprint is under review. When a metamodel is used (for example reduced basis metamodel, but also kriging, regression, ...) for estimating sensitivity indices by Monte Carlo type estimation, a twofold error appears : a sampling error and a metamodel error. Deriving confidence intervals taking into account these two sources of uncertainties is of great interest. We obtained results particularly well fitted for reduced basis metamodels [13]. In a more recent work [69], the authors deal with asymptotic confidence intervals in the double limit where the sample size goes to infinity and the metamodel converges to the true model. Implementations have to be conducted on more general models such as Shallow-Water models. Let us come back to the output of interest. Is it possible to get better error certification when the output is specified. A work in this senses has been submitted, dealing with goal oriented uncertainties assessment [70].

6.4.3.4. Sensitivity analysis with dependent inputs

An important challenge for stochastic sensitivity analysis is to develop methodologies which work for dependent inputs. For the moment, there does not exist conclusive results in that direction. Our aim is to define an analogue of Hoeffding decomposition [82] in the case where input parameters are correlated. A PhD started in October 2010 on this topic (Gaëlle Chastaing). We obtained first results [4], deriving a general functional ANOVA for dependent inputs, allowing defining new variance based sensitivity indices for correlated inputs.

6.4.3.5. Multy-fidelity modeling for risk analysis

Federico Zertuche PhD concerns the modeling and prediction of a digital output from a computer code when multiple levels of fidelity of the code are available. A low-fidelity output can be obtained, for example on a coarse mesh. It is cheaper, but also much less accurate than a high-fidelity output obtained on a fine mesh. In this context, we propose new approaches to relieve some restrictive assumptions of existing methods ([83], [92]) : a new estimating method of the classical cokriging model when designs are not nested and a nonparametric modeling of the relationship between low-fidelity and high-fidelity levels. The PhD takes place in the REDICE consortium and in close link with industry. The first year was also dedicated to the development of a case study in fluid mechanics with CEA in the context of the study of a nuclear reactor.

6.4.4. Multivariate risk indicators

In collaboration with Véronique Maume-Deschamps (ISFA Lyon 1), Elena Di Bernardino (CNAM), Anne-Catherine Favre (LTHE Grenoble) and Peggy Cenac (Université de Bourgogne), we are interested in defining and estimating new multivariate risk indicators. This is a major issue with many applications (environmental, insurance, ...). Two papers were accepted for publication and two other ones are submitted. The first submitted one deals with the estimation of bivariate tails [79]. In [81] and [68] we propose estimation procedures for multivariate risk indicators. In [5] we propose to minimize multivariate risk indicators by using a Kiefer-Wolfowitz approach to the mirror stochastic algorithm.

6.4.5. Quasi-second order analysis for the propagation and characterization of uncertainties in geophysical prediction

We have developed a new approach for the propagation and characterization of uncertainties in geophysical prediction. Most of the method presently used are based on Monte-Carlo type (ensemble) methods, they are expensive from the computational point of view and have received a poor theoretical justification especially in the case of strongly non linear models. We have proposed a new method based on quasi-second order analysis, with a theoretical background and robust for strongly non linear models. Several papers have been published [20], [10], [19], [51] and the application to complex models are presently under development. Igor Gejadze and Victor Shutyaev have been staying both for a total of four weeks in MOISE.

6.5. Image processing

6.5.1. Image processing

Participant: Nicolas Papadakis.

In collaboration with the Inria team MC2 of the Bordeaux-Sud-Ouest center, we investigate the application of image assimilation to medical issues. The objective is here to use MRI images in order to monitor EDP models dealing with tumor growth in lungs or brains. Using such images, we would like to define a patient specific process allowing to calibrate the numerical model with respect to the observed tumor. First works based on convex relaxation of the binary segmentation problem [34] have been realized in this direction by proposing a 3D segmentation method dedicated to glioblastomas from a set of MRI brain images. The obtained automatic segmentation results are very close to specialist manual segmentations (errors of 5%) and will be used as pseudo-observations for an assimilation system based on the numerical model describing the tumor growth. The final issue will be to define an observation operator linking images with the model in order to realize a direct assimilation.

Next, in collaboration with Vicent Caselles (Pompeu Fabra University, Barcelona, Spain) we tackled the problem of histogram equalization of different images. Our aim has been to include spatial information on color repartition during the histogram transfer for inpainting and shadow removal purposes [18]. We also focused with Jean-François Aujol (Institut de Mathématiques de Bordeaux), on the convexification of non linear problems such as optical flow estimation and submitted a jounal paper on this subject in SIAM Journal on Imaging Sciences.

6.5.2. Optimal transport

Participants: Maëlle Nodet, Nicolas Papadakis, Arthur Vidard.

Within the optimal transport project TOMMI funded by the ANR white program, some new algorithms had been proposed to take into account the physics (rigidity, elesticity) of the density to transport [40]. A journal paper has been submitted on this topic in M2AN.

6.6. Mathematical modelling for CFD-environment coupled systems

Participant: Antoine Rousseau.

6.6.1. Minimal-time bioremediation of natural water resources

The objective of this work is to provide efficient strategies for the bioremediation of natural water resources. The originality of the approach is to couple minimal time strategies that are determined on a simplified model with a faithful numerical model for the hydrodynamics. Based on a previous paper that deals with an an implicit representation of the spatial inhomogenity of the resource with a small number of homogeneous compartments (with a system of ODEs), we implement a coupled ODE-PDE system that accounts for the spatial non-homogeneity of pollution in natural resources. The main idea is to implement a Navier-Stokes model in the resource (such as a lake), with boundary conditions that correspond to the output feedback that has been determined to be optimal for the simple ODEs model of a (small) bioreactor. A first mathematical model has been introduced and numerical simulations have been performed in academic situations. During the internship of S. Barbier (co-advised by A. Rousseau and A. Rapaport (INRA-MODEMIC)) we built a reduced model that approximates the reference PDE model thanks to a set of ODEs with parameters. Numerical optimization is performed on these parameters in order to better fit the reference model. This will lead to a publication. In addition, bioremediation algorithms proposed by the authors have been sent to Inria Technology Transfert Services for a patent registration.

6.6.2. Mathematical modelling for the confinement of lagoons

This work deals with the concept of confinement of paralic ecosystems. It is based on a recent paper by E. Frénod that presents a modelling procedure in order to compute the confinement field of a lagoon.
In [9], A. Rousseau and E. Frénod improve the existing model in order to account for tide oscillations in any kind of geometry such as a non-rectangular lagoons with a non-flat bottom. The new model, that relies on PDEs rather than ODEs, is then implemented thanks to the finite element method. Numerical results confirm the feasibility of confinement studies thanks to the introduced model. During the internship of J.-P. Bernard, we implemented the proposed method in a realistic situation, namely the Etang de Thau in Languedoc-Roussillon, France (see Figure 5). This was presented in an international conference [60].

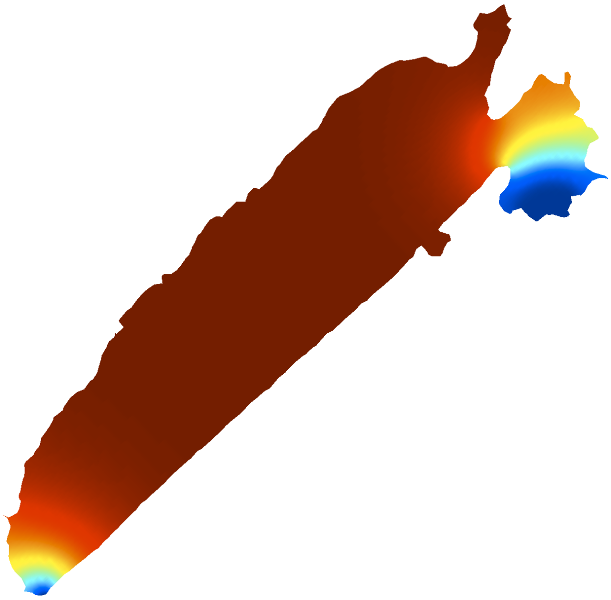


Figure 5. Confinement map in the Thau Lagoon (France). See [60].

6.7. CO_2 Storage

Participant: Céline Helbert.

In collaboration with Bernard Guy (EMSE, Saint-Etienne) and more specifically in the context the PhD of Joharivola Raveloson (EMSE, Saint-Etienne), we are interested in the study of the water-rock interactions in the case of CO_2 storage in geological environment. This work is following the study of Franck Diedro in the same subject [8]. In this study we focus on the scale of observation of geochemical phenomena while taking into account the heterogeneity of the reservoir. This heterogeneity at small and large scale helps to maintain a local variability of the chemical composition and influence reaction rates at the pore as well at the reservoir scale. To connect the parameters at both scale (pore and reservoir) we use deterministic and stochastic simulations of a reactive transport code developed by IFPEN.

6.8. Land Use and Transport models calibration

Participants: Clémentine Prieur, Nicolas Papadakis, Arthur Vidard.

Given the complexity of modern urban areas, designing sustainable policies calls for more than sheer expert knowledge. This is especially true of transport or land use policies, because of the strong interplay between the land use and the transportation systems. Land use and transport integrated (LUTI) modelling offers invaluable analysis tools for planners working on transportation and urban projects. Yet, very few local authorities in charge of planning make use of these strategic models. The explanation lies first in the difficulty to calibrate these models, second in the lack of confidence in their results, which itself stems from the absence of any well-defined validation procedure. Our expertise in such matters will probably be valuable for improving the reliability of these models. To that purpose we participated to the building up of the ANR project CITiES lead by the STEEP EPI. This project has just been accepted and will start early 2013.

MORPHEME Team

5. New Results

5.1. Imaging

5.1.1. ML estimation of wavelet regularization hyperparameters in inverse problems

Participant: Laure Blanc-Féraud.

This work was made in collaboration with Caroline Chaux from LATP (Marseille) and Roberto Cavicchioli and Luca Zanni from University of Modena (Italy).

Parameter estimation, Maximum likelihood estimation, Wavelet transforms, Deconvolution, Gradi- ent meth- ods

We are interested in regularizing hyperparameter estimation by maximum likelihood in inverse problems with wavelet regularization. One parameter per subband is estimated by gradient ascent algorithm.We have to face with two main difficulties: i) sampling the a posteriori image distribution to compute the gradient of the objective function; ii) choosing a suited step-size to ensure good convergence properties of the gradient ascent algorithm. We first show that introducing an auxiliary variable makes the sampling feasible using classical Metropolis-Hastings algorithm and Gibbs sampler. Secondly, we propose an adaptive step-size selection and a line-search strategy to improve the gradient-based method. Good performances of the proposed approach are demonstrated on both synthetic and real data.

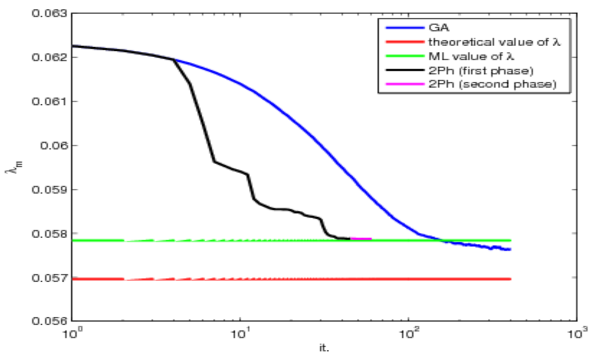


Figure 1. λ_m behavior over iterations of ascent algorithm for a sub band at first level of wavelet decomposition

5.1.2. Joint optimization of noisy image coding and denoising

Participants: Mikael Carlavan, Laure Blanc-Féraud.

this work was made in collaboration with Marc Antonini (I3S), Roberto Camarero and Christophe Latry (CNES) and Yves Bobichon (TAS).

coding, denoising, wavelet transform, global rate-distortion optimization

This work concerns the study of optimal noisy source coding/denoising. A global optimization of the problem is usually difficult to perform as the global fidelity criterion needs to be optimized in the same time over the sets of both coding and denoising parameters. Most of the bibliography in this domain is based on the fact that, for a specific criterion, the global optimization problem can be simply separated into two independent optimization problems: The noisy image should be first optimally denoised and this denoised image should then be optimally coded. In many applications however, the layout of the acquisition imaging chain is fixed and can not be changed, that is a denoising step can not be inserted before coding. For this reason, we are concerned here with the problem of global joint optimization in the case the denoising step is performed, as usual, after coding/decoding. In this configuration, we showed on a simple case how to express the global distortion as a function of the coding and denoising parameters. We presented an algorithm to minimize this distortion to get the optimal values of these parameters. Figure 2 shows results of this joint optimization algorithm, on the classical test image *Barbara*, in comparison to the usual disjoint optimization technique, which consists in selecting the coding and the denoising parameters such that the coding and the denoising errors are independently minimized. On the range of validity of the proposed model, we see that the joint optimized distortion slightly outperforms the disjoint optimized distortion (in the presented example, the PSNR of the reconstructed image increases of $0.4dB$ at 1.85 bits/pixels). The interesting point of the proposed method is that it allows to reach the same global error than the disjoint optimized technique but for a lower coding rate. For example, on this image, the joint optimization technique reaches at 1.42 bits/pixel the same distortion than the one obtained at 2.04 bits/pixels for the disjoint optimization technique. The benefit in terms of compression performances of the joint optimization appears then to be very significant.

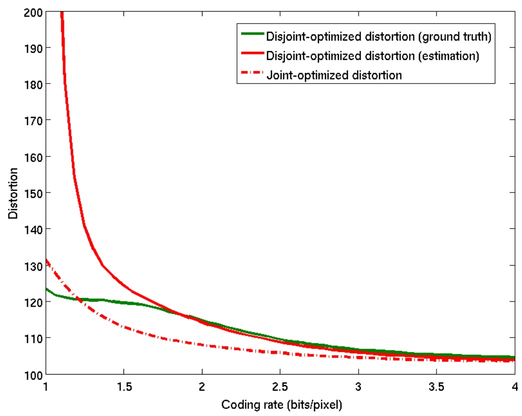


Figure 2. Comparison of the disjoint optimized distortion (ground truth and estimation) to the joint optimized distortion Barbara.

5.1.3. Blind deconvolution

Participants: Saima Ben Hadj, Laure Blanc-Féraud.

This research takes place within the ANR DIAMOND. This work was made in collaboration with Gilles Aubert, Laboratoire J. Dieudonné (CNRS,UNS).

One of our tasks within the ANR Diamond project is the blind restoration of images coming from Confocal laser scanning microscopy (CLSM). CLSM is a powerful technique for studying biological specimens in three dimensions by optical sectioning. Nevertheless, it suffers from some artifacts. First, CLSM images are affected by a depth-variant (DV) blur due to spherical aberrations induced by refractive index mismatch between the different media composing the system as well as the specimen. Second, CLSM images are corrupted with a

Poisson noise due to low illumination. Because of these intrinsic optical limitations, it is essential to remove both DV blur and noise from these images by digital processing.

In this context, we first study space-variant (SV) blur models and prove that a model where the SV point spread function (PSF) is approximated by a convex combination of a set of space-invariant (SI) PSFs is efficient and adequate to the inversion problem [30] [10]. Afterwards, we focus on the non-bind restoration problem and we fit a fast restoration method based on a domain decomposition technique [33] to our DV blur model [10], [9].

Recently, we focus on the blind case. In fact, in practice it is difficult to obtain the DV PSF in spite of the existence of theoretical PSF models [34], because these models are dependent on some unknown acquisition parameters (e.g. the refractive index (RI) of the specimen). Therefore a blind or semi-blind restoration algorithm is needed for this system. We propose two methods for this problem : In the first method, we define a criterion to be jointly minimized w.r.t to the image and the PSF set. In this method, the intensities of each SI PSF are estimated at every voxel. Although the big number of parameters to be estimated, the method allows more freedom on the shape of the PSF which could be more or less deformed according to spherical aberration level. We provide a theoretical proof of the existence of a minimizer of the considered problem [23]. Then, we perform the minimization by following an alternate minimization scheme, each elementary minimization is performed using the recently proposed scaled gradient projection (SGP) algorithm that has shown a fast convergence rate [29]. Results on simulated CLSM images and comparison with another alternate scheme based on a regularized version of the Richardson–Lucy algorithm [31] are shown in Fig. 3 . In the second blind method, we use a Gaussian approximation of each of the SI PSFs. This presents the advantage of significantly reducing the number of parameters to be estimated but constraints the PSF shape. We prove on simulated data that the first method provides more accurate restoration result than the second one.

(a) (b) (c) (d)

Figure 3. (Y, Z) sections of blind restoration results on a simulated CLSM image. (a) original image, (b) simulated observation, (c) restoration using our blind restoration method using SGP algorithm, (d) restoration using a regularized Richardson–Lucy algorithm embedded in an alternate scheme.

5.1.4. Morphogenesis of living organisms

Participant: Grégoire Malandain.

This research takes place within the Inria Large-scale initiative Morphogenetics.

This work was made in collaboration with Christophe Godin and Léo Guignard from Virtual Plants.

super-resolution, SPIM, morphogenesis

We extended a previous work [32] for the reconstruction of microscopic images. In particular, we extended the super-resolution image reconstruction (where several images, acquired from different viewpoints, are fused) to the lightsheet (or SPIM for Selective Plane Illumination Microscope) microscope modality. This modality offers a high acquisition speed, allowing imaging an organism frequently. As an exemple, Phallusia mammillata and Ciona intestinalis embryos can be imaged from 32 cells to around 1000 cells. The organism is captured from four different angles every 2 minutes during 2 hours (collaboration with CRBM Montpellier and EMBL Heidelberg).

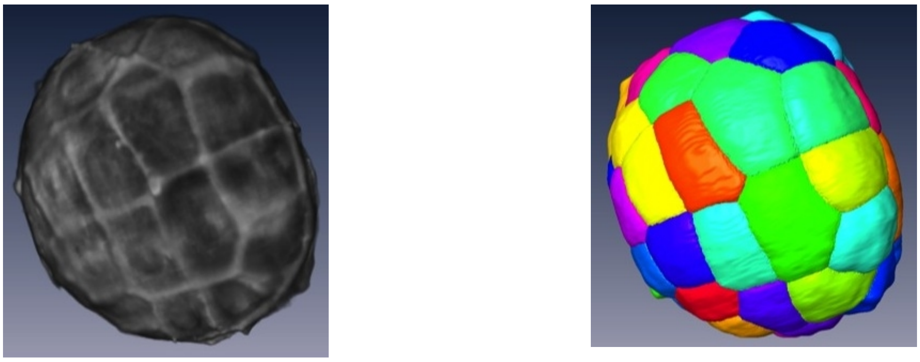


Figure 4. Left: 3D rendering of a reconstructed image of a Phallusia mammillata embryo; right: segmentation of cells.

5.2. Features Extraction

5.2.1. Axon extraction from fluorescent confocal microscopy images

Participants: Alejandro Mottini, Xavier Descombes, Florence Besse.

It it known that the analysis of axonal topologies allows biologists to study the causes of neurological diseases such as Fragile X Syndrome and Spinal Muscular Atrophy. In order to perform the morphological analysis of axons, it is first necessary to segment them. Therefore, the automatic extraction of axons is a key problem in the field of neuron axon analysis.

For this purpose, biologists label single neurons within intact adult Drosophila fly brains and acquire 3D fluorescent confocal microscopy images of their axonal trees. These images need to be segmented.

In our work presented in [16], we propose a new approach for the automatic extraction of axons from fluorescent confocal microscopy images which combines algorithms for filament enhancement, binarization, skeletonization and gap filling in a pipeline capable of extracting the axons containing a single labeled neuron. Unlike other segmentation methods found in the literature, the proposed is fully automatic and designed to work on 3D image stacks. This allows us to analyze large image databases.

The method performance was tested on 12 real 3D images and the results quantitatively evaluated by calculating the RMSE between the tracing done by an experienced biologist and the automatic tracing obtained by our method. The good results obtained in the validation show the potential use of this technique in helping biologists for extracting axonal trees from confocal microscope images (see figures 5 and 6).

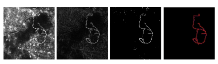


Figure 5. Results obtained on each step of the algorithm for one image stack (maximum intensity projections). From left to right: original image, filament enhancement, binarization and final result.

5.2.2. Dendrite spine detection from X-ray tomographic volumes

Participants: Anny Hank, Xavier Descombes, Grégoire Malandain.

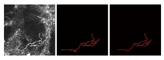


Figure 6. Comparison between original image (left), our result (middle) and ground truth (right) for two images (maximum intensity projections).

We have developped an automated algorithm for detecting dendritic spines from XRMT data. XRMT data allows imaging a large volume of tissue, and therefore a higher number of spines than laser scanning microscopy. We have shown that despite the lower image quality compared to microscopic data, we were able to extract dendritic spines. The main idea of the proposed approach is to define a mask for performing the spine detection without facing the false alarms problem as we introduce some information on spines localization. We therefore first extract the dendrites themselves and then compute the spine mask based on prior knowledge on their distance to dendrites. To extract dendrite we first compute the medial axis thanks to a multi-scale Hessian-based method. Then, we extract segments by a 3D Hough transform and reconstruct the dendrites using a conditional dilation. The spine mask is defined nerby the detected dendrites using anatomical parameters described in the literature. A point process defined on this mask provides the spine detection.

To exemplify the proposed approach, a subvolume ($220 \times 180 \times 100$) has been extracted from a XRMT volume that is given on figure 7 . As expected, the spines appear as small objects, whose size is close to the image resolution, along the tubular structures representing dendrites. Using the localization information to detect spine is essential to prevent false alarms due to noise or to the deviation of dendrites from a cylinder model. Figure 7 shows the detected dendrite medial axis and the obtained spine detection. The obtained results are promising and correspond to a visual inspection of the data. Forthcoming validation study will allow to better assess the quality of the detection by providing a quantitative evaluation.

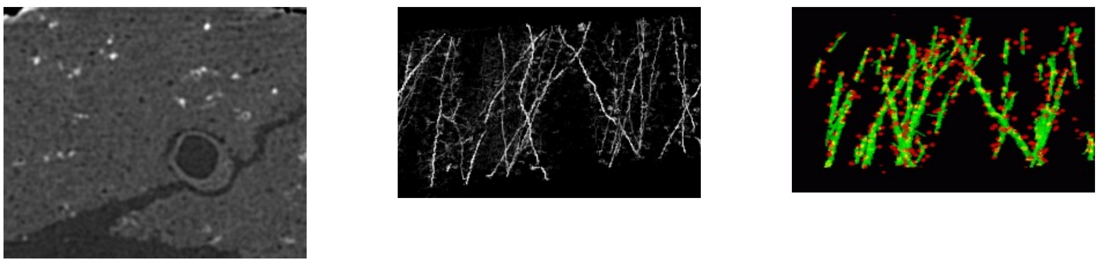


Figure 7. XMRT slice (left), dendrites medial axis (middle) and spine detection (in red) (right)

5.2.3. Cell detection

Participant: Xavier Descombes.

This work was done in collaboration with Emmanuel Soubies and Pierre Weiss from ITAV (Toulouse)

We have proposed some improvements of the Multiple Birth and Cut algorithm (MBC) in order to extract nuclei in 2D and 3D images. We have introduced a new contrast invariant energy that is robust to degradations encountered in fluorescence microscopy (e.g. local radiometry attenuations). Another contribution of this work is a fast algorithm to determine whether two ellipses (2D) or ellipsoids (3D) intersect. Finally, we propose a new heuristic that strongly improves the convergence rates. The algorithm alternates between two birth steps. The first one consists in generating objects uniformly at random and the second one consists in perturbing the current configuration locally. Performance of this modified birth step is evaluated and examples on various image types show the wide applicability of the method in the field of bio-imaging.

Figure 8 left shows the segmentation result on a Drosophila embryo obtained using SPIM imaging. This is a rather easy case, since nuclei shapes vary little. The images are impaired by various defects: blur, stripes and attenuation. Despite this relatively poor image quality, the segmentation results are almost perfect. The computing time is 5 minutes using a C++ implementation. The image size is 700 × 350. Figure 8 right presents a more difficult case, where the image is highly deteriorated. Nuclei cannot be identified in the image center. Moreover, nuclei variability is important meaning that the state space size χ is large. Some nuclei are in mitosis (see e.g. top-left). In spite of these difficulties, the MBC algorithm provides acceptable results. They would allow to make statistics on the cell location and orientation, which is a major problem in biology. The computing times for this example is 30 minutes.

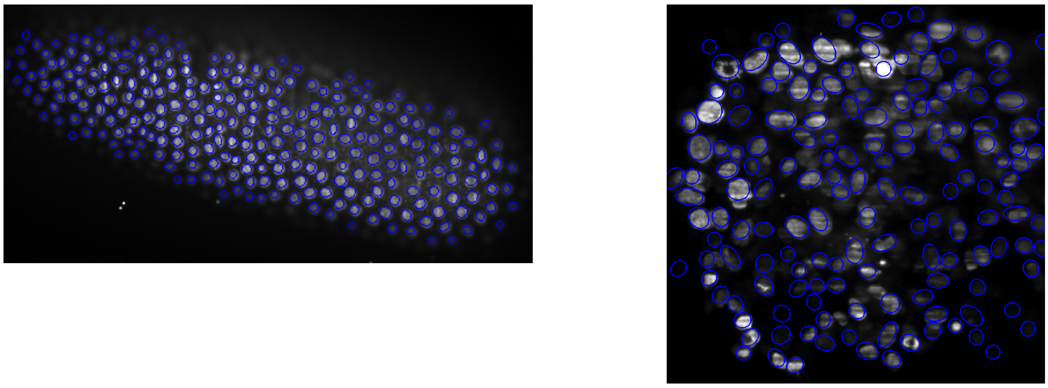


Figure 8. 2D segmentations of a nuclei of Drosophila embryo (left) and a multicellular tumor spheroid (right).

5.2.4. Spermatozoid tracking

Participants: Clarens Caraccio, Xavier Descombes.

In this work, we have proposed an algorithm for tracking spermatozoid in a sequence of confocal images. We first detect the spermatozoids by thresholding the result of a top hat operator. The thresold is automatically estimated using Otsu's method. We then analyse the different connected components to detect overlaps between adjacent spermatozoids. Temporal neighbors are selected based on the spatial consistency of the object sets between two consecutive time. A first result is given on figure 9 .

5.3. Classification

5.3.1. Axon morphology comparison using elastic shape analysis

Participants: Alejandro Mottini, Xavier Descombes, Florence Besse.

It is known that neuronal morphology impacts network connectivity, thus providing information on its functioning. Moreover, it allows the characterization of pathological states. Therefore, the analysis of the morphological differences between normal and pathological structures is of paramount importance.

We present a new method for comparing reconstructions of axonal trees (obtained, for example, by applying our segmentation method on confocal microscopy images of normal and mutant axonal trees) which takes into account both topological and geometrical information and is based on the Elastic Shape Analysis Framework. The method computes the geodesic between two axons in a space of tree like shapes, and the distance between the two is defined as the length of the geodesic. Moreover, our method is capable of showing how one axon transforms into the other by taking intermediate points in the geodesic.

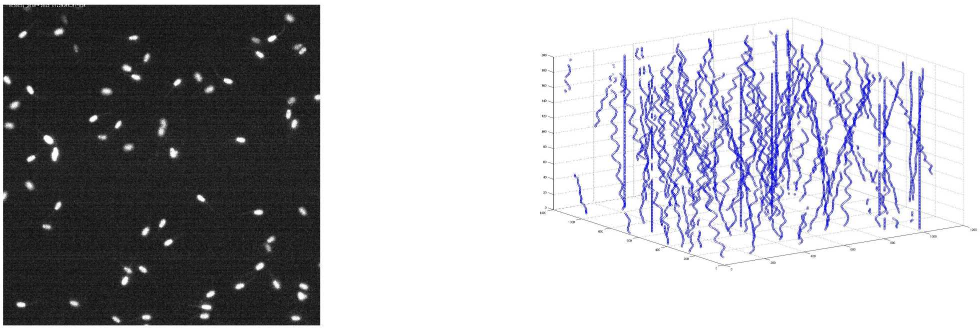


Figure 9. Original confocal image and estimated spermatozoid trajectories.

We consider two axonal trees T_1 and T_2, each consisting of an axon and several branches (and possibly sub branches). All are represented by 3D open curves in $\mathbb{R}^3$ (see Figure 10). We start by defining the matching function M such that $M : (0, 1, 2, ...n) \times (0, 1, ..., m)$, where n and m are the number of branches in T_1 and T_2 respectively. The matching function matches the branches of the two trees, for example, by assigning branch $i = 1$ of T_1 to branch $j = 3$ of T_2. We then define a branch function C which indicates, for a given time t_c, how many branches remain after $\beta(t_c)$ (see Figure 10). We only take into account branches which have a match in the other axonal tree. Finally, we define the distance between two axonal trees T_1, T_2 as:

$$D(T_1, T_2) = \min_M d((\beta_1(t), C1(t, M)), (\beta_2(t), C2(t, M))) + \sum_{(i,j)} \alpha_{i,j} M(i, j) D(T_1(i), T_2(j)) \tag{31}$$

where β_k is the main curve (axon) of tree k, C_k its branch function, M the matching function, $\alpha_{i,j}$ a weight parameter and $D(T_1(i), T_2(j))$ the distance between the matched branches of the two trees. All distances between simple curves are calculated using the elastic shape analysis framework.

The method performance was tested on a group of 22 (11 normal and 11 mutant) 3D images, each containing one axonal tree manually segmented by an experienced biologist from a set of real confocal microscopy images. The mean and standard deviation of the inter and intra class distances between the neurons were calculated and results suggest that the proposed method is able to distinguish between the two populations (an average interpopulation to intrapopulation distance ratio of 1:21 and 1:28 were obtained). In addition, we computed the optimum transformations between axons. An example is shown in figure 11 . This result was obtained by taking intermediate points along the geodesic between the two trees.

Figure 10. Axonal tree diagrams (a) and their corresponding C functions for a given M (b).

5.3.2. Vascular network segmentation from X-ray tomographic volumes

Participant: Xavier Descombes.

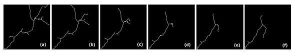


Figure 11. Optimum transformation between two axonal trees (transformation starts in (a) and finishes in (f), maximum intensity projections).

This work was made in collaboration with Franck Plouraboué and Abdelakim El Boustani from IMFT, Caroline Fonta from CerCo, Géraldine LeDuc from ESRF, Raphael Serduc from INSERM and Tim Weitkamp from Synchrotron Soleil.

Micro-tomography produces high resolution images of biological structures such as vascular networks. We have defined a new approach for segmenting vascular network into pathological and normal regions from considering their micro-vessel 3D structure only. We consider a partition of the volume obtained by a watershed algorithm based on the distance from the nearest vessel. Each territory, defined as Local Vascular Territory (*a Local Vascular Territory (LVT) is a connected region corresponding to the catchment bassin associated with a vascular element. It can be obtained through the watershed computation on the opposite distance map from the vessels and is not connected to the sample border.*), is characterized by its volume and the local vascular density. The volume and density maps are first regularized by minimizing the total variation, within a Markov Random Field framework, using a graph cut algorithm . Then, a new approach is proposed to segment the volume from the two previous restored images using an iterative algorithm based on hypothesis testing. We consider the variables density and volume for each LVT and the populations constituted by the different classes obtained by the segmentation at a given step. Classes which are not statistically significantly different are merged using a MANOVA. This blind segmentation provides different regions which have been interprated by expert as tumor, necrosis, tumor periphery and sane tissue 12 .

5.3.3. Statistical analysis of skin pigmentation under treatment

Participants: Sylvain Prigent, Xavier Descombes.

This work was partially funded by a contract with Galderma R&D [http://www.galderma.com/RampD.aspx]. It was made in collaboration with J. Zerubia from Ayin team.

multispectral imaging, skin, hyperpigmentation, hypothesis tests, statistical inferences

One of the steps to evaluate the efficacy of a therapeutic solution is to test it on a clinical trial involving several populations of patients. Each population receives a studied treatment and a reference treatment for the disease. For facial hyper-pigmentation, a group of N_e patients receives the treatment on one cheek and a comparator on the other. The comparator can be a reference treatment or a placebo. To this end patients are selected to have the same hyper-pigmentation severity on the two cheeks. Then multi-spectral images are taken at different time t along the treatment period.

We propose a methodology to assess the efficacy a treatment by calculating three differential criteria: the darkness, the area and the homogeneity. The darkness measure the average intensity of the disease on a gray scaled image I obtained by a linear combination of the spectral bands of the original multi-spectral image. A differential darkness is then obtained by measuring the deviation between the initial measurement at time t_0, and the measurement at time t_k. The differential area criterion is calculated by analyzing the histogram of $I_{diff} = I_{t_0} - I_{t_k}$ a difference gray scale image between two measurements in a time series. The differential homogeneity criterion is obtaining with a multi-scale analysis of I_{diff} adapted from the Statistical Parametric Mapping (SPM) methodology. Indeed, statistical inferences allow to assign a probability of change to each region of I_{diff} above a set of thresholds. These probabilities are calculated with respect to the maximum intensity and the spatial extend of each region. An integration of the obtained statistical map denoted SM, allows to get a homogeneity criterion.

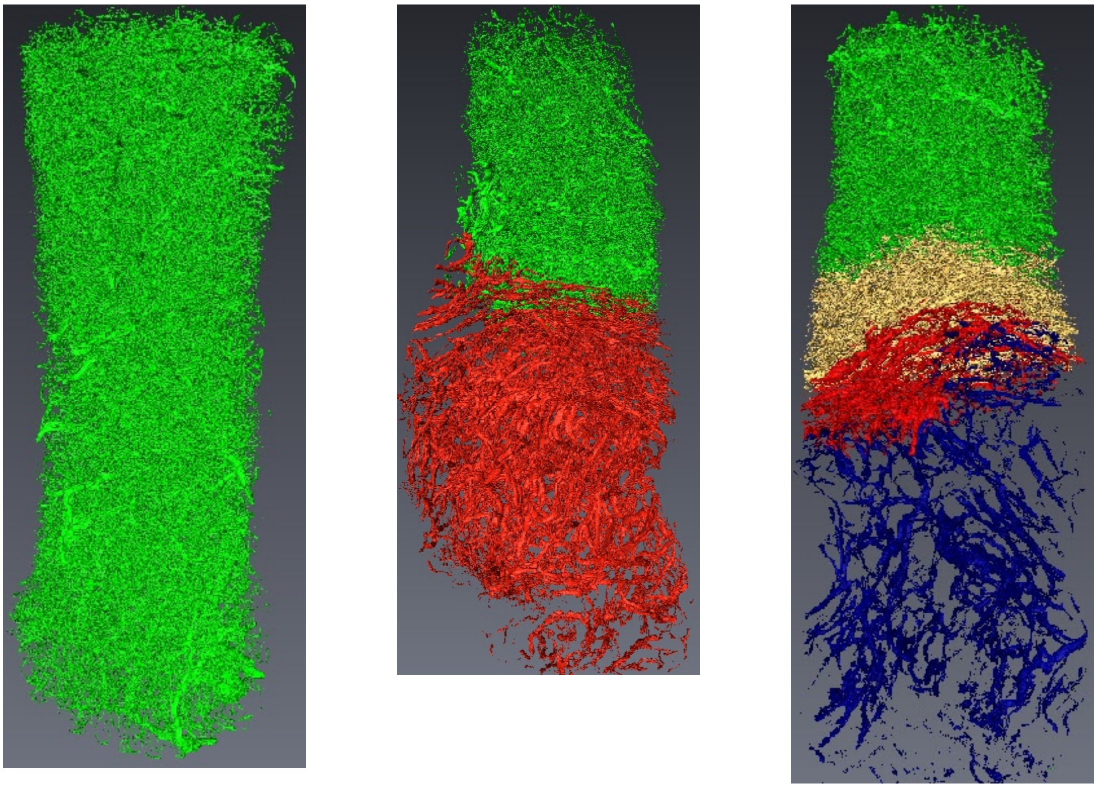


Figure 12. Examples of segmentation: tumor (red), necrosis (blue), tumor periphery (yellow) and sane (green)

The figure 13 illustrates the differential score calculated on a patient whose pathology decreases during the clinical trial. The proposed differential score have been tested in a full clinical study and provided results that agreed with the clinical analysis. This work have been patented and published in Inria research reports [25], [26].

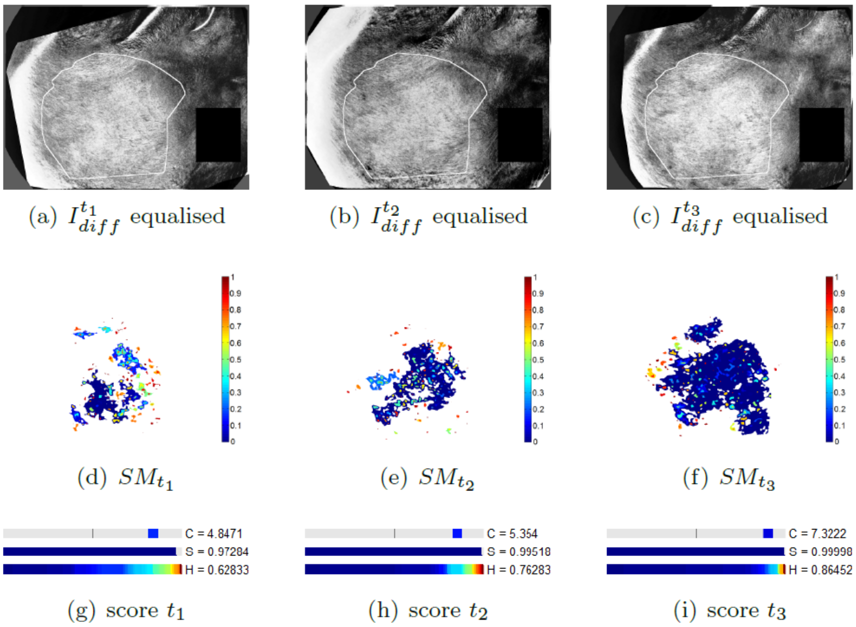


Figure 13. I_{diff}, SM and differential score for the three measurements t_1, t_2, t_3 calculated for a patient whose disease decreases.

5.3.4. A Recursive Approach For Multiclass Support Vector Machine: Application to automatic classification of endomicroscopic videos

Participants: Alexis Zubiolo, Eric Debreuve.

This work is made in collaboration with Barbara André (Mauna Kea Technologies)

The problem of automatic image (or video, or object) classification is to find a function that maps an image to a class or category among a number of predefined classes. An image can be viewed as a vector of high-dimension. In practice, it is preferable to deal with a synthetic signature of lower dimension. Therefore, the two classical steps of image classification are: image signature extraction and signature-based image classification. The classification rule can be learned from a set of training sample images manually classified by experts. This is known as supervised statistical learning where *statistical* refers to the use of samples and *supervised* refers to the sample classes being provided. We are interested in the learning aspect of the multiclass [1] problem when using a binary classification approach as a building block. We chose the Support Vector Machine (SVM), a well-known binary classifier.

Among the proposed extensions of binary classification methods to multiclass (three classes or more), the one-versus-one and one-versus-all approaches are the most popular ones. Let us suppose that there are $p \geq 3$ classes. The idea of the one-versus-all strategy is to oppose to any of the classes the union of the remaining $p - 1$ classes. Then, p SVM classifiers are determined, each one scoring, say, positively for one of the classes.

[1] Traditionally in classification, *multiclass* means "three classes or more" while the two-class case is referred to as binary classification.

The one-versus-one strategy opposes the classes by pair for all possible pairs. Therefore, $\frac{p(p-1)}{2}$ SVMs are determined and classification is performed by a majority vote.

As an alternative to these aforementioned strategies (as well as to other, less popular ones), we developed a recursive learning strategy. A tree of SVMs is built, achieving three goals: a fair balance in the number of samples used in each binary SVM learnings, a logarithmic complexity for classification ($\log_2(p)$ compared to the linear or quadratic complexities of one-versus-all or one-versus-one, respectively), and a coherent, incremental classification procedure (as opposed to selecting the final class based on possibly competing partial decisions). During learning, at each node of the tree, a combinatorial search is performed to determine an optimal separation of the current classification problem into two sub-problems. The proposed method was applied to automatic classification of endomicroscopic videos.

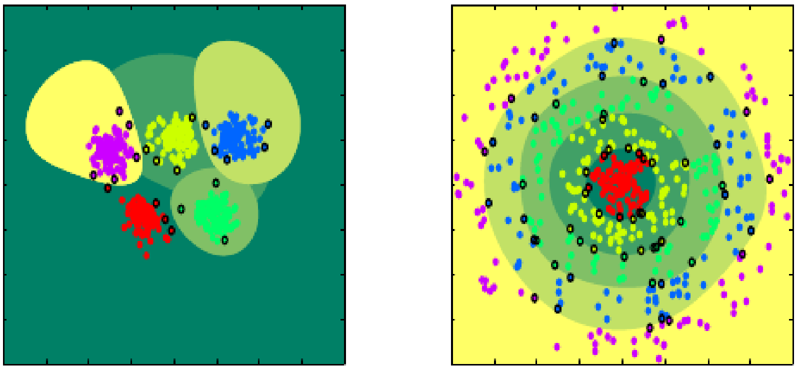


Figure 14. Illustration of the proposed recursive approach for multiclass Support Vector Machine. Colored dots: learning feature samples; Encircled dots: computed support vectors; Colored areas: computed class regions. Left: classical example; Right: concentric example.

5.4. Modeling

5.4.1. Tracking Growing Axons in Fluorescent Microscopy Images

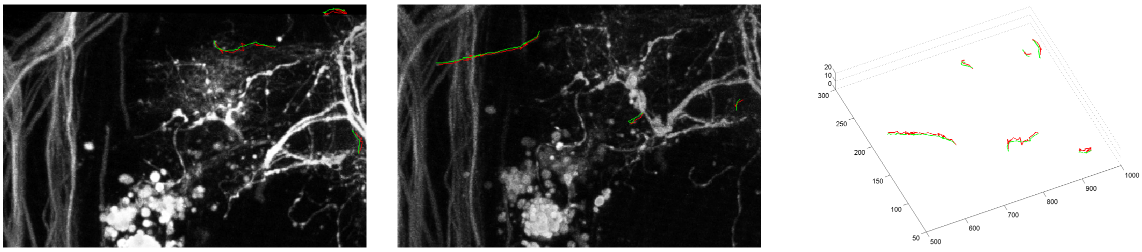
Participants: Huei Fang Yang, Florence Besse, Xavier Descombes.

This work has been done in collaboration with Caroline Medioni from iBV.

Analyzing how growing axons correctly reach their target neurons is essential for biologists to better understand the development of a nervous system. Analysis of the properties of axon growth requires detecting axonal tips and tracking their trajectories within complex and large data sets. When performed manually, the tracking task is arduous and time-consuming. To this end, we proposed a tracking method, based on the particle filtering technique, to follow the traces of axonal tips that appear as small bright spots in the 3D $+ t$ fluorescent two-photon microscopy images exhibiting low signal-to-noise ratios (SNR) and complex background. Our tracking method uses multiple dynamic models in the proposal distribution to predict the positions of the growing axons. Moreover, it incorporates object appearance, motion characteristics of the growing axons, and filament information in the computation of the observation model. The integration of these three sources results in improved accuracy of recovered trajectories. The experimental results obtained from the microscopy images, presented in Figure 15 , showed that the proposed method can successfully estimate trajectories of growing axons, demonstrating its effectiveness even under the presence of noise and complex background.

5.4.2. Trajectory Simulation of Growing Axons:

Participants: Huei Fang Yang, Florence Besse, Xavier Descombes.



(a) (b) (c)

Figure 15. Visual comparison between the tracking results of the proposed method and the manually created ground truth in 2D and 3D. The red trajectories are produced by the proposed method, and the green are the ground truth manually created by the expert; both are overlaid on the MIPs (a) and (b) and visualized in 3D (c). The computer generated tracks are consistent with the ground truth in general, with minor differences between the estimated positions and the ground truth positions. The differences are caused by the noise and by the effect of complex background.

This work has been done in collaboration with Caroline Medioni from iBV.

It is established in biology that axons reach their target cells in the developing nervous system by the guidance of molecular gradients. To better understand how growing axons react to the molecular cues, either attractant or repellent, we simulated the trajectories of growing axons using a mathematical model that investigates the effect of molecular gradients on the axon's growth angle. Figure 16 shows the simulated trajectories of 50 growing axons. The initial position of axons is $(0, 0)$, and the red point on the right denotes the target cell that secretes the attractant cue.

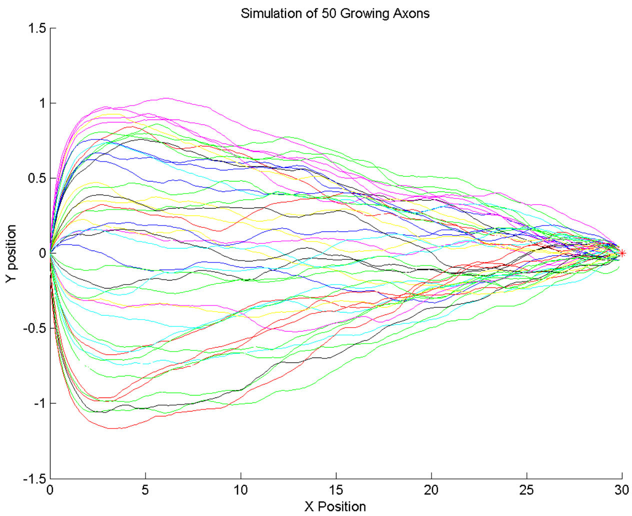


Figure 16. Simulated trajectories of 50 growing axons. The initial position of 50 axons is set to $(0, 0)$, and the red point on the right represents the target cell that secretes the attractant cue.

NEUROMATHCOMP Project-Team

5. New Results

5.1. Neural Networks as dynamical systems

5.1.1. Dynamics and spike trains statistics in conductance-based Integrate-and-Fire neural networks with chemical and electric synapses

Participants: Rodrigo Cofré, Bruno Cessac [correspondent].

We investigate the effect of electric synapses (gap junctions) on collective neuronal dynamics and spike statistics in a conductance-based Integrate-and-Fire neural network, driven by a Brownian noise, where conductances depend upon spike history. We compute explicitly the time evolution operator and show that, given the spike-history of the network and the membrane potentials at a given time, the further dynamical evolution can be written in a closed form. We show that spike train statistics is described by a Gibbs distribution whose potential can be approximated with an explicit formula, when the noise is weak. This potential form encompasses existing models for spike trains statistics analysis such as maximum entropy models or Generalized Linear Models (GLM). We also discuss the different types of correlations: those induced by a shared stimulus and those induced by neurons interactions. This work has been presented in several conferences [43], [45], [46], [47], [31], [48] and submitted to Chaos, Solitons and Fractals [13].

5.1.2. Parameter estimation in spiking neural networks: a reverse-engineering approach

Participants: Horacio Rostro-Gonzalez [Holistic Electronics Research Lab, University of Cyprus], Bruno Cessac [correspondent], Thierry Viéville [Inria Mnemosyne].

This work presents a reverse engineering approach for parameter estimation in spiking neural networks (SNNs). We consider the deterministic evolution of a time-discretized network with spiking neurons, where synaptic transmission has delays, modeled as a neural network of the generalized integrate and fire type. Our approach aims at by-passing the fact that the parameter estimation in SNN results in a non-deterministic polynomial-time hard problem when delays are to be considered. Here, this assumption has been reformulated as a linear programming (LP) problem in order to perform the solution in a polynomial time. Besides, the LP problem formulation makes explicit the fact that the reverse engineering of a neural network can be performed from the observation of the spike times. Furthermore, we point out how the LP adjustment mechanism is local to each neuron and has the same structure as a 'Hebbian' rule. Finally, we present a generalization of this approach to the design of input–output (I/O) transformations as a practical method to 'program' a spiking network, i.e. find a set of parameters allowing us to exactly reproduce the network output, given an input. Numerical verifications and illustrations are provided. This work has been published in Journal of Neural Engineering [24].

5.2. Mean field approaches

5.2.1. Noise-induced behaviors in neural mean field dynamics

Participants: Olivier Faugeras [correspondent], Geoffroy Hermann, Jonathan Touboul [Inria Bang].

The collective behavior of cortical neurons is strongly affected by the presence of noise at the level of individual cells. In order to study these phenomena in large-scale assemblies of neurons, we consider networks of firing-rate neurons with linear intrinsic dynamics and nonlinear coupling, belonging to a few types of cell populations and receiving noisy currents. Asymptotic equations as the number of neurons tends to infinity (mean field equations) are rigorously derived based on a probabilistic approach. These equations are implicit on the probability distribution of the solutions which generally makes their direct analysis difficult. However, in our case, the solutions are Gaussian, and their moments satisfy a closed system of nonlinear ordinary differential equations (ODEs), which are much easier to study than the original stochastic network equations, and the statistics of the empirical process uniformly converge towards the solutions of these ODEs. Based on this description, we analytically and numerically study the influence of noise on the collective behaviors, and compare these asymptotic regimes to simulations of the network. We observe that the mean field equations provide an accurate description of the solutions of the network equations for network sizes as small as a few hundreds of neurons. In particular, we observe that the level of noise in the system qualitatively modifies its collective behavior, producing for instance synchronized oscillations of the whole network, desynchronization of oscillating regimes, and stabilization or destabilization of stationary solutions. These results shed a new light on the role of noise in shaping collective dynamics of neurons, and gives us clues for understanding similar phenomena observed in biological networks. This work has been published in the SIAM Journal on Applied dynamical Systems [25].

5.2.2. Mean-field description and propagation of chaos in networks of Hodgkin-Huxley neurons

Participants: Javier Baladron, Diego Fasoli, Olivier Faugeras [correspondent], Jonathan Touboul [Inria Bang].

We derive the mean-field equations arising as the limit of a network of interacting spiking neurons, as the number of neurons goes to infinity. The neurons belong to a fixed number of populations and are represented either by the Hodgkin-Huxley model or by one of its simplified version, the FitzHugh-Nagumo model. The synapses between neurons are either electrical or chemical. The network is assumed to be fully connected. The maximum conductances vary randomly. Under the condition that all neurons' initial conditions are drawn independently from the same law that depends only on the population they belong to, we prove that a propagation of chaos phenomenon takes place, namely that in the mean-field limit, any finite number of neurons become independent and, within each population, have the same probability distribution. This probability distribution is a solution of a set of implicit equations, either nonlinear stochastic differential equations resembling the McKean-Vlasov equations or non-local partial differential equations resembling the McKean-Vlasov-Fokker-Planck equations. We prove the well posedness of the McKean-Vlasov equations, i.e. the existence and uniqueness of a solution. We also show the results of some numerical experiments that indicate that the mean-field equations are a good representation of the mean activity of a finite size network, even for modest sizes. These experiments also indicate that the McKean-Vlasov-Fokker-Planck equations may be a good way to understand the mean-field dynamics through, e.g. a bifurcation analysis. This work has appeared in the Journal of Mathematical Neuroscience [11].

5.3. Neural fields theory

5.3.1. Localized radial bumps of a neural field equation on the Euclidean plane and the Poincaré disk

Participants: Grégory Faye [correspondent], James Rankin, David, J.B. Lloyd [Department of Mathematics at the University of Surrey].

We analyze radially symmetric localized bump solutions of an integro-differential neural field equation posed in Euclidean and hyperbolic geometry. The connectivity function and the nonlinear firing rate function are chosen such that radial spatial dynamics can be considered. Using integral transforms, we derive a PDE for the neural field equation in both geometries and then prove the existence of small amplitude radially symmetric spots bifurcating from the trivial state. Numerical continuation is then used to path-follow the spots and their bifurcations away from onset in parameter space. It is found that the radial bumps in Euclidean geometry are linearly stable in a larger parameter region than bumps in the hyperbolic geometry. We also find and path-follow localized structures that bifurcate from branches of radially symmetric solutions with D6-symmetry and D8-symmetry in the Euclidean and hyperbolic cases, respectively. Finally, we discuss the applications of our results in the context of neural field models of short term memory and edges and textures selectivity in a hypercolumn of the visual cortex. This work has been accepted for publication in Nonlinearity [57].

5.3.2. Center manifold for delayed neural fields equations

Participants: Olivier Faugeras [correspondent], Romain Veltz [Salk Institute,San Diego,USA].

We develop a framework for the study of delayed neural fields equations and prove a center manifold theorem for these equations. Specific properties of delayed neural fields equations make it impossible to apply existing methods from the literature concerning center manifold results for functional differential equations. Our approach for the proof of the center manifold theorem uses the original combination of results from Vanderbauwhede and colleagues together with a theory of linear functional differential equations in a history space larger than the commonly used set of time-continuous functions. This work has been submitted to the SIAM Journal on Applied Mathematics and is under revision [27].

5.3.3. Reduction method for localized solutions

Participant: Grégory Faye [correspondent].

We present a reduction method to study localized solutions of an integrodifferential equation defined on the Poincaré disk. This equation arises in a problem of texture perception modeling in the visual cortex. We first derive a partial differential equation which is equivalent to the initial integrodifferential equation and then deduce that localized solutions which are radially symmetric satisfy a fourth order ordinary differential equation. This work has appeared in the “Comptes Rendus Mathematique” [15].

5.3.4. Spatially localized solutions

Participants: Pascal Chossat, Grégory Faye [School of Mathematics,University of Minnesota, correspondent], James Rankin.

The existence of spatially localized solutions in neural networks is an important topic in neuroscience as these solutions are considered to characterize working (short-term) memory. We work with an unbounded neural network represented by the neural field equation with smooth firing rate function and a wizard hat spatial connectivity. Noting that stationary solutions of our neural field equation are equivalent to homoclinic orbits in a related fourth order ordinary differential equation, we apply normal form theory for a reversible Hopf bifurcation to prove the existence of localized solutions; further, we present results concerning their stability. Numerical continuation is used to compute branches of localized solution that exhibit snaking-type behaviour. We describe in terms of three parameters the exact regions for which localized solutions persist. This work has appeared in the Journal of Mathematical Biology [16].

5.3.5. Bumps in the Poincaré disk

Participants: Grégory Faye [School of Mathematics,University of Minnesota, correspondent], David, J.B. Loyd, James Rankin.

We analyze radially symmetric localized bump solutions of an integro-differential neural field equation posed in Euclidean and hyperbolic geometry. The connectivity function and the nonlinear firing rate function are chosen such that radial spatial dynamics can be applied. Using integral transforms, we derive a PDE of the neural field equation in both geometries and then prove the existence of small amplitude radially symmetric spots bifurcating from the trivial state. Numerical continuation is then used to path-follow the spots and their bifurcations away from onset in parameter space. It is found that the radial bumps in Euclidean geometry are linearly stable in a larger parameter region than bumps in the hyperbolic geometry. We also find and path follow localized structures that bifurcate from branches of radially symmetric solutions with D6-symmetry and D8-symmetry in the Euclidean and hyperbolic cases, respectively. Finally, we discuss the applications of our results in the context of neural field models of short term memory and edges and textures selectivity in a hypercolumn of the visual cortex. This work has been submitted to Nonlinearity.

5.4. Spike trains statistics

5.4.1. Natural image identification from spike train analysis

Participants: Geoffrey Portelli, Olivier Marre [Institution de la Vision, Paris, France], Marc Antonini [Laboratoire I3S, UMR CNRS, Universite´ Nice Sophia Antipolis, France], Michael Berry II [Princeton Neuroscience Institute, Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA], Pierre Kornprobst [correspondent].

We started a new activity to analyse how natural images are encoded in retinal output. This work is related to [67], [72] where synthetic stimuli are used. Here, we recorded a population of 100-200 ganglion cells of a salamander retina, while flashing 720 natural images from the Torralba database [71] plus one control image, each repeated 10 times. We characterized the response of each cell by two parameters : the latency of the first spike after the stimulus onset, and the firing rate. A distribution of these two features was then estimated for each neuron and natural image. Pooling the information across all the neurons, a discriminability coefficient between pairs of image is proposed, using either the rate or the latency, or both. We also provide a way to identify a given image among others based on the rate–latency distributions. Preliminary results have been presented in [40]. Results showed that, on average, the discriminability was better based on the latency than on the rate. The most discriminable pairs were different using the rate or the latency, so these two features conveyed complementary information. In addition, we observe a similar evolution of the identification performance when the rate, or the latency, or both are used.

5.4.2. Spike train statistics from empirical facts to theory: the case of the retina

Participants: Bruno Cessac [correspondent], Adrian Palacios [CINV-Centro Interdisciplinario de Neurociencia de Valparaiso, Universidad de Valparaiso].

This work focuses on methods from statistical physics and probability theory allowing the analysis of spike trains in neural networks. Taking as an example the retina we present recent works attempting to understand how retina ganglion cells encode the information transmitted to the visual cortex via the optical nerve, by analyzing their spike train statistics. We compare the maximal entropy models used in the literature of retina spike train analysis to rigorous results establishing the exact form of spike train statistics in conductance-based Integrate-and-Fire neural networks. This work has been published in Mathematical Problems in Computational Biology and Biomedicine, F. Cazals and P. Kornprobst, Springer [55].

5.4.3. Gibbs distribution analysis of temporal correlations structure in retina ganglion cells

Participants: Juan-Carlos Vasquez, Olivier Marre [Institution de la Vision, Paris, France], Adrian Palacios [CINV-Centro Interdisciplinario de Neurociencia de Valparaiso, Universidad de Valparaiso], Michael Berry II [Princeton Neuroscience Institute, Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA], Bruno Cessac [correspondent].

We present a method to estimate Gibbs distributions with spatio-temporal constraints on spike trains statistics. We apply this method to spike trains recorded from ganglion cells of the salamander retina, in response to natural movies. Our analysis, restricted to a few neurons, performs more accurately than pairwise synchronization models (Ising) or the 1-time step Markov models (Marre et al. (2009)) to describe the statistics of spatio-temporal spike patterns and emphasizes the role of higher order spatio-temporal interactions. This work has been presented in several conferences [29], [30], [28] and pusblished in J. Physiol. Paris [26].

5.4.4. Spatio-temporal spike trains analysis for large scale networks using maximum entropy principle and Monte-Carlo method

Participants: Hassan Nasser, Olivier Marre [Institut de la Vision, Paris, France], Bruno Cessac [correspondent].

Understanding the dynamics of neural networks is a major challenge in experimental neuroscience. For that purpose, a modelling of the recorded activity that reproduces the main statistics of the data is required. We present a review on recent results dealing with spike train statistics analysis using maximum entropy models (MaxEnt). Most of these studies have been focusing on modelling synchronous spike patterns, leaving aside the temporal dynamics of the neural activity. However, the maximum entropy principle can be generalized to the temporal case, leading to Markovian models where memory effects and time correlations in the dynamics are properly taken into account. We also present a new method based on Monte-Carlo sampling which is suited for the fitting of large-scale spatio-temporal MaxEnt models. The formalism and the tools presented here will be essential to fit MaxEnt spatio-temporal models to large neural ensembles. This work has been presented in several conferences [54], [51], [53], [52] and accepted in Journal of Statistical Mechanics [22].

5.4.5. Spike train statistics and Gibbs distributions

Participants: Rodrigo Cofré, Bruno Cessac [correspondent].

We introduce Gibbs distribution in a general setting, including non stationary dynamics, and present then three examples of such Gibbs distributions, in the context of neural networks spike train statistics: (i) Maximum entropy model with spatio-temporal constraints; (ii) Generalized Linear Models; (iii) Conductance based Inte- grate and Fire model with chemical synapses and gap junctions. This leads us to argue that Gibbs distributions might be canonical models for spike train statistics analysis. This work has been presented in several conferences [43], [31] and submitted to J. Physiol. Paris [12].

5.5. Visual Neuroscience

5.5.1. Neural fields models for motion integration: Characterising the dynamics of multi-stable visual motion stimuli

Participants: Olivier Faugeras, Pierre Kornprobst, Guillaume S. Masson [Institut de Neurosciences de la Timone, UMR 6193, CNRS, Marseille, France], Andrew Meso [Institut de Neurosciences de la Timone, UMR 6193, CNRS, Marseille, France], James Rankin.

We are investigating the temporal dynamics of the neural processing of a multi-stable visual motion stimulus with two complementary approaches: psychophysical experiments and mathematical modelling. The so called “barber pole” stimulus is considered with an aperture configuration that supports horizontal (H), diagonal (D) or vertical (V) perceived directions for the same input. The phenomenon demonstrates an interesting variable and dynamic competition for perceptual dominance between underlying neural representations of the three directions. We study the temporal dynamics of this phenomenon with a neural fields, population-level representation of activity in MT, a cortical area dedicated to motion estimation. Numerical tools from bifurcation analysis are used to investigate the model’s behaviour in the presence of different types of input; this general approach could be applied to a range of neural fields models that are typically studied only in terms of their spontaneous activity. The model reproduces known multistable behaviour in terms of the predominant interpretations (percepts) of the barber pole stimulus.

We probe the early processing from stimulus presentation to initial perceived direction (before perceptual reversals). The basic dynamic properties of the early transition from D to H/V are well predicted by the model. This work has been presented in the European Conference on Vision Perception (ECVP) [38], [41] and it has been published in [23].

We are extending this work to investigate the longer term dynamics for which perceptual reversals are known to occur, due to competition between 1D motion cues aligned with the grating’s motion direction and 2D motion cues aligned with aperture edges. This work has been presented in the Vision Sciences Society 12th Annual Meeting (VSS) [39], [42].

5.6. Neuromorphic Vision

Participants: Khaled Masmoudi [Laboratoire I3S, UMR CNRS, Universite´ Nice Sophia Antipolis, France], Marc Antonini [Laboratoire I3S, UMR CNRS, Universite´ Nice Sophia Antipolis, France], Pierre Kornprobst.

In the scope of Khaled Masmoudi's PhD [9], we have developed bio-inspired schemes for image coding. This is a new area of research on which very few teams are committed. We have proposed schemes for encoding/decoding images directly using the functional architecture of the retina and the properties of its spiking output (e.g., using Laplacian pyramids model and the rank-order coding [69], [68][20], [21] and the Virtual Retina simulator [37] from Adrien Wohrer during his PhD [79], [78]).

NUMED Project-Team

6. New Results

6.1. New result 1

Numed has developped a general strategy and generic softwares (to be released soon) to allow populational parametrization on complex models like PDEs.

PARIETAL Project-Team

6. New Results

6.1. Randomized cluster-based predictive model

Participants: Gaël Varoquaux [Correspondant], Bertrand Thirion, Alexandre Gramfort.

Functional neuroimaging can measure the brain's response to an external stimulus. It is used to perform brain mapping: identifying from these observations the brain regions involved. This problem can be cast into a linear supervised learning task where the neuroimaging data are used as predictors for the stimulus. Brain mapping is then seen as a support recovery problem. On functional MRI (fMRI) data, this problem is particularly challenging as i) the number of samples is small due to limited acquisition time and ii) the variables are strongly correlated. We propose to overcome these difficulties using sparse regression models over new variables obtained by clustering of the original variables. The use of randomization techniques, e.g. bootstrap samples, and clustering of the variables improves the recovery properties of sparse methods. We demonstrate the benefit of our approach on an extensive simulation study as well as two fMRI datasets.

More details can be found in [32].

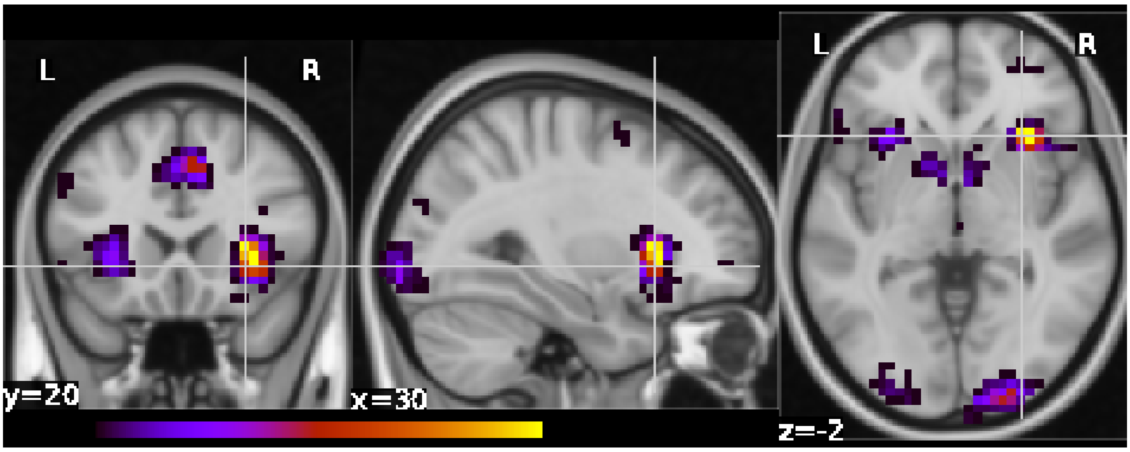


Figure 3. The randomized cluster-based predictive model can be used to predict the behavior of the subject, such as the gain in a gambling task [33]. More importantly, the support of the resulting classifier is indeed sparse and provides a reliable definition of the truly involved regions.

6.2. Random Projections for Outlier Detection

Participants: Gaël Varoquaux, Bertrand Thirion, Jean-Baptiste Poline, Virgile Fritsch [Correspondant].

Medical imaging datasets often contain deviant observations, the so-called outliers, due to acquisition or preprocessing artifacts or resulting from large intrinsic inter-subject variability. These can undermine the statistical procedures used in group studies as the latter assume that the cohorts are composed of homogeneous samples with anatomical or functional features clustered around a central mode. The effects of outlying subjects can be mitigated by detecting and removing them with explicit statistical control. With the emergence of large medical imaging databases, exhaustive data screening is no longer possible, and automated outlier detection methods are currently gaining interest. The datasets used in medical imaging are often high-dimensional and strongly correlated. The outlier detection procedure should therefore rely on high-dimensional statistical multivariate models. However, state-of-the-art procedures are not well-suited for such high-dimensional settings. In this work, we introduce regularization in the Minimum Covariance Determinant framework and investigate different regularization schemes. We carry out extensive simulations to provide backing for practical choices in absence of ground truth knowledge. We demonstrate on functional neuroimaging datasets that outlier detection can be performed with small sample sizes and improves group studies.

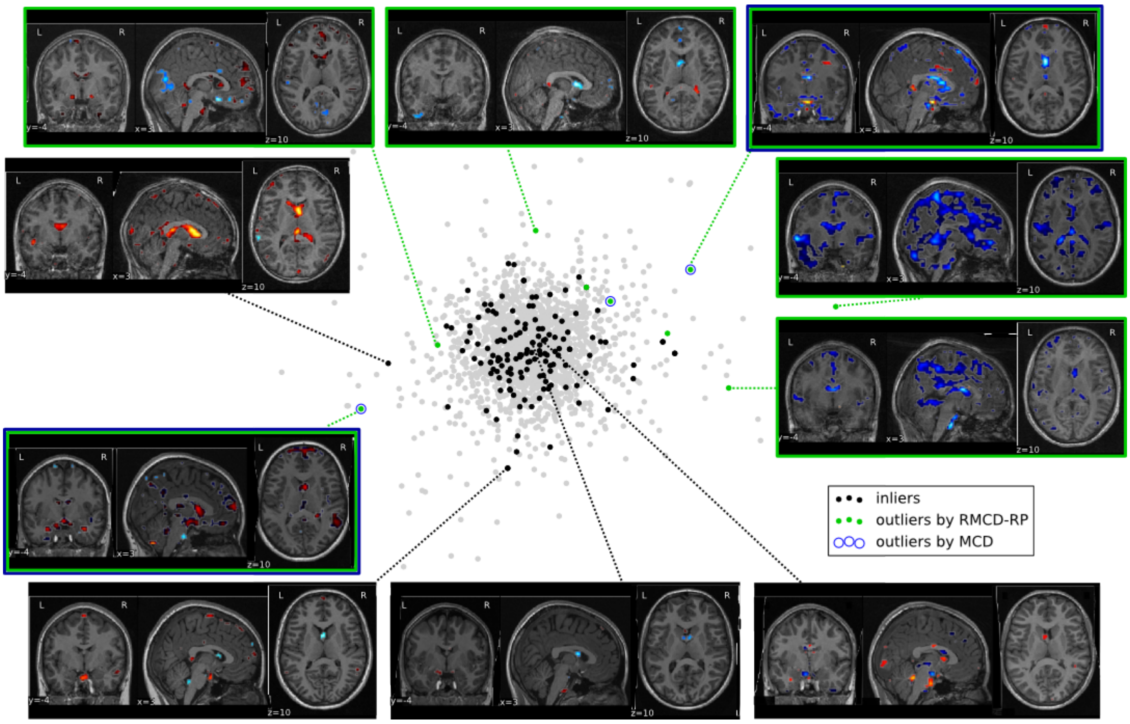


Figure 4. A large set of images can be mined for structures using the regularized MCD framework, which reveals both standard and unusual patterns in these images.

More details can be found in [11].

6.3. Registration of brain images based on Currents

Participants: Pierre Fillard, Bertrand Thirion, Viviana Siless [correspondant].

We present an extension of the diffeomorphic Geometric Demons algorithm which combines the iconic registration with geometric constraints. Our algorithm works in the log-domain space, so that one can efficiently compute the deformation field of the geometry. We represent the shape of objects of interest in the space of currents which is sensitive to both location and geometric structure of objects. Currents provide a distance between geometric structures that can be defined without specifying explicit point-to-point correspondences. We demonstrate this framework by registering simultaneously T1 images and 65 fiber bundles consistently extracted in 12 subjects and compare it against non-linear T1, tensor, and multi-modal T1+ Fractional Anisotropy (FA) registration algorithms. Results show the superiority of the Log-domain Geometric Demons over their purely iconic counterparts.

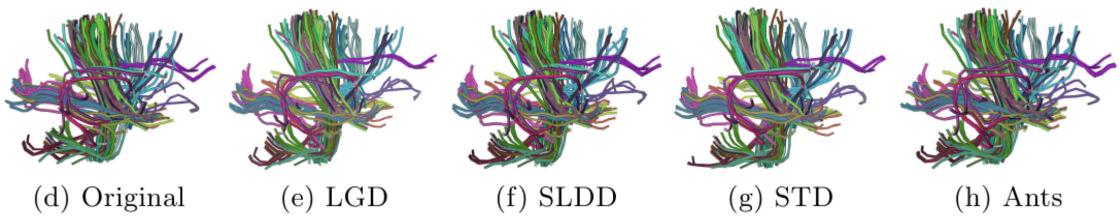


Figure 5. Comparison of the fiber registration through various algorithms. We display a moving and a reference fiber for 29 selected bundles. The proposed approach (SLDD) outperforms state-of-the art alternatives that do not take into account the fiber geometry explicitly.

More details can be found in [31].

6.4. Structured Sparsity for brain mapping

Participants: Gaël Varoquaux [Correspondant], Bertrand Thirion, Alexandre Gramfort.

Reverse inference, or *brain reading*, is a recent paradigm for analyzing functional magnetic resonance imaging (fMRI) data, based on pattern recognition and statistical learning. This approach aims at decoding brain activity by predicting some cognitive variables related to brain activation maps. Reverse inference takes into account the multivariate information between voxels and is currently the only way to assess how precisely some cognitive information is encoded by the activity of neural populations within the whole brain. However, it relies on a prediction function that is plagued by the curse of dimensionality, since there are far more features than samples, i.e., more voxels than fMRI volumes. To address this problem, different methods have been proposed, such as, among others, univariate feature selection, feature agglomeration and regularization techniques. In this work, we consider a sparse hierarchical structured regularization. Specifically, the penalization we use is constructed from a tree that is obtained by spatially-constrained agglomerative clustering. This approach encodes the spatial structure of the data at different scales into the regularization, which makes the overall prediction procedure more robust to inter-subject variability. The regularization used induces the selection of spatially coherent predictive brain regions simultaneously at different scales. We test our algorithm on real data acquired to study the mental representation of objects, and we show that the proposed algorithm not only delineates meaningful brain regions but yields as well better prediction accuracy than reference methods.

More details can be found in [15].

6.5. A Novel Sparse Graphical Approach for Multimodal Brain Connectivity Inference

Participants: Bertrand Thirion, Jean-Baptiste Poline, Gaël Varoquaux [Correspondant], Bernard Ng.

Despite the clear potential benefits of combining fMRI and diffusion MRI in learning the neural pathways that underlie brain functions, little methodological progress has been made in this direction. In this work, we propose a novel multimodal integration approach based on sparse Gaussian graphical model for estimating brain connectivity. Casting functional connectivity estimation as a sparse inverse covariance learning problem, we adapt the level of sparse penalization on each connection based on its anatomical capacity for functional interactions. Functional connections with little anatomical support are thus more heavily penalized. For validation, we showed on real data collected from a cohort of 60 subjects that additionally modeling anatomical capacity significantly increases subject consistency in the detected connection patterns. Moreover, we demonstrated that incorporating a connectivity prior learned with our multimodal connectivity estimation approach improves activation detection.

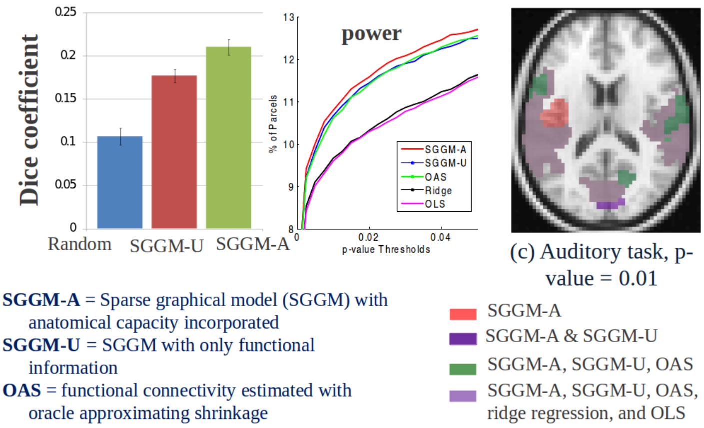


Figure 6. The information conveyed by anatomical connectivity improves the estimation of functional connectivity, as it makes it more reproducible. It also enhances the power of fMRI activation detection studies when used as a prior on these activation maps.

More details can be found in [26].

6.6. Transfer learning for met-analyses of functional neuroimaging datasets

Participants: Bertrand Thirion, Jean-Baptiste Poline, Gaël Varoquaux, Yannick Schwartz [Correspondant].

Typical cohorts in brain imaging studies are not large enough for systematic testing of all the information contained in the images. To build testable working hypotheses, investigators thus rely on analysis of previous work, sometimes formalized in a so-called meta-analysis. In brain imaging, this approach underlies the specification of regions of interest (ROIs) that are usually selected on the basis of the coordinates of previously detected effects. In this work, we propose to use a database of images, rather than coordinates, and frame the problem as transfer learning: learning a discriminant model on a reference task to apply it to a different but related new task. To facilitate statistical analysis of small cohorts, we use a sparse discriminant model that selects predictive voxels on the reference task and thus provides a principled procedure to define ROIs. The benefits of our approach are twofold. First it uses the reference database for prediction, i.e. to provide potential biomarkers in a clinical setting. Second it increases statistical power on the new task. We demonstrate on a set of 18 pairs of functional MRI experimental conditions that our approach gives good prediction. In addition, on a specific transfer situation involving different scanners at different locations, we show that voxel selection based on transfer learning leads to higher detection power on small cohorts.

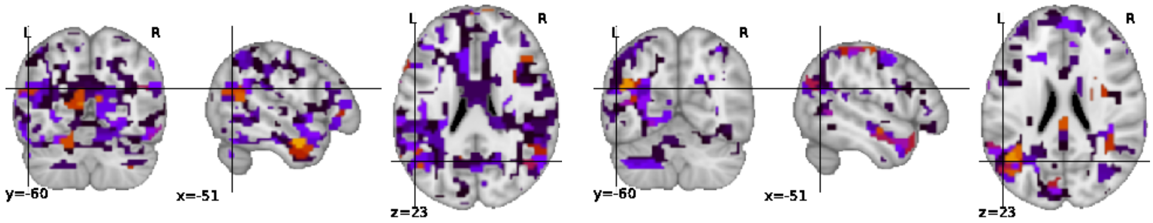


Figure 7. The brain regions that reliably predict that the subject is listening to Korean versus native (french) language (left) are similar to those that can be used to predict that the subject is listening an unintelligible language (jabberwoky) as opposed to their native (french) language (right).

More details can be found in [29] and [30].

6.7. Learning to rank medical images

Participants: Bertrand Thirion, Gaël Varoquaux, Alexandre Gramfort, Fabian Pedregosa [Correspondant].

Medical images can be used to predict a clinical score coding for the severity of a disease, a pain level or the complexity of a cognitive task. In all these cases, the predicted variable has a natural order. While a standard classifier discards this information, we would like to take it into account in order to improve prediction performance. A standard linear regression does model such information, however the linearity assumption is likely not be satisfied when predicting from pixel intensities in an image. In this work we address these modeling challenges with a supervised learning procedure where the model aims to order or rank images. We use a linear model for its robustness in high dimension and its possible interpretation. We show on simulations and two fMRI datasets that this approach is able to predict the correct ordering on pairs of images, yielding higher prediction accuracy than standard regression and multi-class classification techniques.

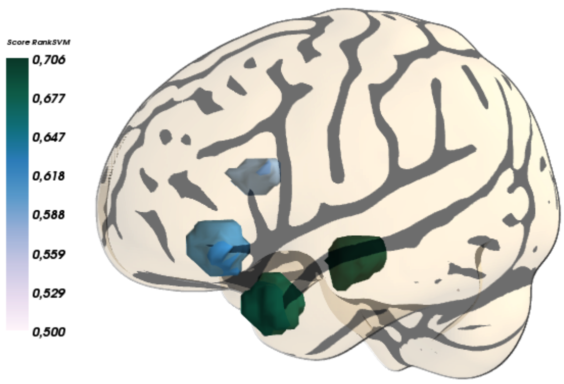


Figure 8. Based on a ranking procedure, the information present in different regions of interest of the brain volume can be used to predict a cognitive feature, in that case the level of complexity of sentences heared by the subject.

More details can be found in [27] and [28].

6.8. Decoding four letter words from brain activations

Participants: Bertrand Thirion, Alexandre Gramfort [Correspondant].

Word reading involves multiple cognitive processes. To infer which word is being visualized, the brain first processes the visual percept, deciphers the letters, bigrams, and activates different words based on context or prior expectation like word frequency. In this contribution, we use supervised machine learning techniques to decode the first step of this processing stream using functional Magnetic Resonance Images (fMRI). We build a decoder that predicts the visual percept formed by four letter words, allowing us to identify words that were not present in the training data. To do so, we cast the learning problem as multiple classification problems after describing words with multiple binary attributes. This work goes beyond the identification or reconstruction of single letters or simple geometrical shapes and addresses a challenging estimation problem, that is the prediction of multiple variables from a single observation, hence facing the problem of learning multiple predictors from correlated inputs.

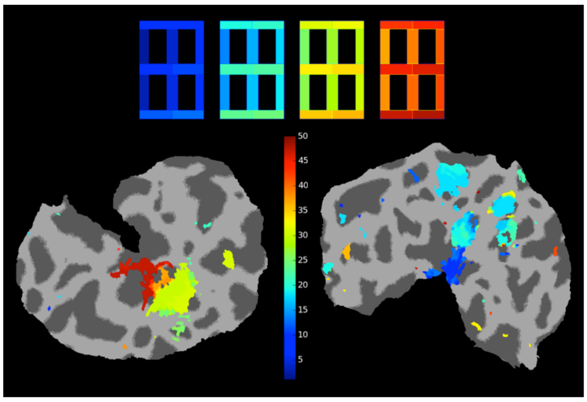


Figure 9. The bars of a word presented on a fixed visual brain activate specific domains of the visual field, and thus can be decoded through this marked. This makes it possible to identify a four letters word presented on a screen.

More details can be found in [22].

POMDAPI Project-Team (section vide)

REO Project-Team

6. New Results

6.1. Mathematical and numerical analysis of fluid-structure interaction problems

Participants: Cristóbal Bertoglio Beltran, Muriel Boulakia, Miguel Ángel Fernández Varela, Sébastien Martin, Jean-Frédéric Gerbeau, Jimmy Mullaert, Marina Vidrascu.

- In [26], we study a three-dimensional fluid-structure interaction problem. The motion of the fluid is modeled by the Navier-Stokes equations and we consider for the elastic structure a finite-dimensional approximation of the equation of linear elasticity. The time variation of the fluid domain is not known a priori, so we deal with a free boundary value problem. Our main result yields the local in time existence and uniqueness of strong solutions for this system.
- In [31], a robust finite volume method for the solution of high-speed compressible flows in multi-material domains involving arbitrary equations of state and large density jumps is presented. One of the main contributions of this paper is a tabulation method based on a sparsegrid approximation to solve very efficiently two-phase Riemann problems for arbitrary equations of state. The proposed computational method is illustrated with the three-dimensional simulation of the dynamics of an underwater explosion bubble.
- In [52] we analyze the performances of several Luenberger observers to estimate the state of a fluid-structure interaction model for hemodynamics, when the measurements are assumed to be restricted to displacements or velocities in the solid. The present framework establishes that these methods are very attractive strategies (compared, e.g., to classical variational techniques) to perform state estimation.
- In [51] we analyze two 3D-0D coupling approaches in which a fractional-step projection scheme is used in the fluid. We introduce and analyze an implicitly 3D-0D coupled formulation with enhanced stability properties and which requires a negligible additional computational cost. The theoretical stability results are confirmed by meaningful numerical experiments in patient specific geometries coming from medical imaging.
- In [55] we introduce a class of explicit Robin-Neumann schemes for the explicit coupling of a general thin-structure (e.g., viscoelastic and non-linear) with an incompressible fluid. These methods generalize the displacement correction schemes introduced in [32]. A priori stability and convergence error estimates show that optimal first-order accuracy can be achieved with appropriate extrapolation and without compromising stability. A deep numerical study confirms the theoretical findings.
- In [64] we present two-dimensional simulations of chemotactic self-propelled bacteria swimming in a viscous fluid. Self-propulsion is modelled by a couple of forces of same intensity and opposite direction applied on the rigid bacterial body and on an associated region in the fluid representing the flagellar bundle. The orientations of the individual bacteria are subjected to random changes, with a frequency that depends on the surrounding oxygen concentration, in order to favor the direction of the concentration gradient.
- In [40] we propose a method of modeling sail structures which captures the wrinkling behavior of such structures. The method is validated through experimental and analytical test cases, particularly in terms of wrinkling prediction. An enhanced wrinkling index is proposed as a valuable measure characterizing the global wrinkling development on the deformed structure. The method is based on a pseudo-dynamic finite element procedure involving non-linear MITC shell elements. The major advantage compared to membrane models generally used for this type of analysis is that no ad hoc

wrinkling model is required to control the stability of the structure. We demonstrate our approach to analyse the behavior of various structures with spherical and cylindrical shapes, characteristic of downwind sails over a rather wide range of shape constitutive parameters. In all cases convergence is reached and the overall flying shape is most adequately represented, which shows that our approach is a most valuable alternative to standard techniques to provide deeper insight into the physical behaviour. Limitations appear only in some very special instances in which local wrinkling-related instabilities are extremely high and would require specific additional treatments.

6.2. Numerical methods for fluid mechanics and application to blood flows

Participants: Grégory Arbia, Jean-Frédéric Gerbeau, Sébastien Martin, Saverio Smaldone, Marc Thiriet, Irène Vignon-Clementel.

- In [18], a procedure for modeling the heart valves is presented. Instead of modeling complete leaflet motion, leaflets are modeled in open and closed configurations. This method enables significant computational savings compared to complete fluid-structure interaction and contact modeling, while maintaining realistic three-dimensional velocity and pressure distributions near the valve, which is not possible from lumped parameter modeling. To illustrate the versatility of the model, realistic and patient-specific simulations are presented, as well as comparison with complete fluid-structure interaction simulation.
- [37] paves the way for a complete patient-specific fluid-structure vascular modeling in which all types of available measurements could be used to estimate uncertain parameters of biophysical and clinical relevance. We propose a complete methodological chain for the identification of the parameters involved in a model for external tissue support of blood vessels, using patient image data. We demonstrate the use of this framework in a realistic application case involving hemodynamics in the thoracic aorta. The estimation of the boundary support parameters proves successful, in particular in that direct modeling simulations based on the estimated parameters are more accurate than with a previous manual expert calibration.
- In [27] we study the image-based blood flow in the first generation of the pulmonary arterial tree. This patient-specific study is aimed at assessing effects of lung deformation and vascular resistance on the pulmonary blood flow, especially during the acute phase of a pneumothorax and after recovery. Arterial geometry was extracted up to the fifth generation from computed tomography images, and reconstructed. An unsteady laminar flow with a given set of resistances at outlets was modeled. Adaptation is set to match perfusion to ventilation.
- In [44], [36] we study the reciprocal effect of blood circulation and high-intensity focused ultrasound on the temperature field in the liver. High-intensity focused ultrasound (HIFU) is used as a thermal ablation process to eliminate tumors in different body's organs. Blood flow has a cooling effect. Conversely, ultrasounds are responsible for acoustic streaming. A three-dimensional acoustics-thermal-fluid coupling model is carried out to compute the temperature field a given hepatic cancerous region.
- The use of elaborate closed-loop lumped parameter network (LPN) models of the heart and the circulatory system as boundary conditions for 3D simulations can provide valuable global dynamic information, particularly for patient specific simulations. In [30], we have developed and tested a numerical method to couple a 3D Navier-Stokes finite-element formulation and a reduced model of the rest of the circulation, keeping the coupling robust but modular. For Neumann boundaries, implicit, semi-implicit, and explicit quasi-Newton formulations are compared within the time-implicit coupling scheme. The requirements for coupling Dirichlet boundary conditions are also discussed and compared to that of the Neumann coupled boundaries. Both these works were key for applications where blood flows in different directions during the cardiac cycle and where coupling with the rest of the circulation is instrumental (see the shunt optimization application [29]).

- Boundary conditions in patient-specific blood flow simulations is key because pressure and flow within the modeled domain are driven by the interplay between the local 3D hemodynamics and the rest of the circulation. However, these boundary conditions are rarely the measured variables. In [45], we showed how one can go from patient-specific clinical data (MRI and catheterization) to simulation input parameters, including modeling assumptions and the impact of both on simulation results. We explained how Windkessel models and more involved LPN can be calibrated.

  In [34], we developed two multi-scale models, each including the 3D model of the surgical junction constructed from MRI, and a closed-loop LPN derived from pre-operative data obtained from two patients prior to Stage 2 Fontan palliation of single ventricle congenital heart disease. "Virtual" surgeries were performed and a corresponding multi-scale simulation predicted the patient's post-operative hemodynamic conditions, tested under different physiological conditions. The impact of the surgical junction geometry on the global circulation was contrasted with variations of key physiological parameters.
- In [19], a similar 3D multiscale model was used but for the Stage 3 Fontan palliation. Several studies have been done to optimize the geometry of the surgical connection, to minimizing energy losses and improving surgical outcomes, but usually without taking into account respiration or exercise. A respiration model that modulates the extravascular pressures in the thoracic and abdominal cavities was implemented. Results showed that the preoperative model is able to realistically capture cardiac and respiratory oscillations compared to the venous Doppler velocity tracings. Three virtual surgical alternatives were coupled to the LPN and then investigated under rest and exercise conditions.
- In [29], such a 3D-closed loop LPN model was integrated with an automated derivative-free optimization algorithm in an idealized systemic-to-pulmonary shunt anatomy (Stage 1 Fontan palliation). The goal was to optimize shunt geometries. Clinicians selected three objective functions to be maximized: (1) systemic, (2) coronary, and (3) combined systemic and coronary oxygen. Results showed the geometries associated with the favored delivery, the origin of coronary artery flow being driven by the shunt position as well. The results made only sense when the 3D domain was connected to a closed-loop model of the circulation.
- A novel Y-shaped baffle was proposed for the Stage 3 Fontan operation achieving overall superior hemodynamic performance compared with traditional designs. Previously, we investigated if and how the inferior vena cava flow (which contains an important biological hepatic factor) could be best distributed among both lungs. In [41] we proposed a multi-step method for patient-specific optimization of such surgeries to study the effects of boundary conditions and geometry on hepatic factor distribution (HFD). The resulting optimal Y-graft geometry largely depended on the patient left/right pulmonary flow split. Unequal branch size and constrained optimization on energy efficiency were explored. Two patient-specific examples showed that optimization-derived Y-grafts effectively improved HFD.

6.3. Numerical methods for cardiac electrophysiology

Participants: Muriel Boulakia, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Vincent Martin, Elisa Schenone.

- In [62], we propose a surface-based electrophysiology formulation, motivated by the modeling of thin structures such as cardiac atria, which greatly reduces the size of the computational models. Our model is specifically devised to retain the key features associated with the anisotropy in the diffusion effects induced by the fiber architecture, with rapid variations across the thickness which cannot be adequately represented by naive averaging strategies. The model relies on a detailed asymptotic analysis in which we identify a limit model and establish strong convergence results. We also provide detailed numerical assessments which confirm an excellent accuracy of the surface-based model – compared with the reference 3D model – including in the representation of a complex phenomenon, namely, spiral waves.

6.4. Lung and respiration modeling

Participants: Laurent Boudin, Paul Cazeaux, Bérénice Grec, Muriel Boulakia, Anne-Claire Egloffe, Benoit Fabreges, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Céline Grandmont, Stéphane Liwarek, Sébastien Martin, Ayman Moussa.

- [59], [60]:We are concerned here with identifiability, stability properties and estimates for the inverse problem of identifying a Robin coefficient on some non accessible part of the boundary from available data on the other part of boundary corresponding to solutions of the Stokes equations. In [59], we first consider a steady state two-dimensional Stokes problem and study the identifiability of Robin coefficient and then we establish a stability estimate of logarithm type using a global Carleman inequality. We then consider the unsteady problem. In [60]:We prove hölderian and logarithmic stability estimates associated to the unique continuation property for the Stokes system. The proof of these results is based on local Carleman inequalities. In the second part, these estimates on the fluid velocity and on the fluid pressure are applied to solve the inverse problem of identifying a Robin coefficient. For this identification parameter problem, we obtain a logarithmic stability estimate under the assumption that the velocity of a given reference solution stays far from 0 on a part of the boundary where Robin conditions are prescribed.
- In [61] we are interested in the mathematical modeling of the propagation of sound waves in the lung parenchyma, which is a foam–like elastic material containing millions of air– filled alveoli. In this study, the parenchyma is governed by the linearized elasticity equations and the air by the acoustic wave equations. The geometric arrangement of the alveoli is assumed to be periodic with a small period $\varepsilon > 0$. We consider the time–harmonic regime forced by vibrations induced by volumic forces. We use the two–scale convergence theory to study the asymptotic behavior as ε goes to zero and prove the convergence of the solutions of the coupled fluid–structure problem to the solution of a linear–elasticity boundary value problem.
- In [53] we develop and study numerically a model to describe some aspects of sound propagation in the human lung, considered as a deformable and viscoelastic porous medium (the parenchyma) with millions of alveoli filled with air. Transmission of sound through the lung above 1 kHz is known to be highly frequency–dependent. We pursue the key idea that the viscoelastic parenchyma structure is highly heterogeneous on the small scale ε and use two–scale homogenization techniques to derive effective acoustic equations for asymptotically small ε. This process turns out to introduce new memory effects. The effective material parameters are determined from the solution of frequency–dependent micro–structure cell problems. We propose a numerical approach to investigate the sound propagation in the homogenized parenchyma using a Discontinuous Galerkin formulation. Numerical examples are presented.
- In [22], we consider the Maxwell-Stefan model of diffusion previously introduced. We provide a qualitative and quantitative mathematical and basic numerical analysis of the model.
- In [65] we propose an integrated model for oxygen transfer into the blood, coupled with a lumped mechanical model for the ventilation process. We aim at investigating oxygen transfer into the blood at rest or exercise. The first task consists in describing nonlinear effects of the oxygen transfer under normal conditions. We also include the possible diffusion limitation in oxygen transfer observed in extreme regimes involving parameters such as alveolar and venous blood oxygen partial pressures, capillary volume, diffusing capacity of the membrane, oxygen binding by hemoglobin and transit time of the red blood cells in the capillaries. The second task consists in discussing the oxygen concentration heterogeneity along the path length in the acinus.
- In [43] we presented preliminary work on a multiscale 3D-0D airflow model to study differences between healthy and emphysema rats. The 0D model parameters were estimated from experimental data. 3D Navier-Stokes simulations were performed in healthy lungs, and in homogenous and heterogeneous emphysema lungs.

6.5. Miscellaneous

Participants: Laurent Boudin, Jean-Frédéric Gerbeau, Damiano Lombardi, Sébastien Martin, Marina Vidrascu, Irène Vignon-Clementel.

- In [56], a reduced-order model algorithm, based on approximations of Lax pairs, is proposed to solve nonlinear evolution partial differential equations. Contrary to other reduced-order methods, like Proper Orthogonal Decomposition, the space where the solution is searched for evolves according to a dynamics specific to the problem. It is therefore well-suited to solving problems with progressive waves or front propagation. Numerical examples are shown for the KdV and FKPP (nonlinear reaction diffusion) equations, in one and two dimensions.
- In [21], we investigate the asymptotic behaviour of the solutions to the non-reactive fully elastic Boltzmann equations for mixtures in the diffusive scaling. We deal with cross sections such as hard spheres or cut-off power law potentials. We use Hilbert expansions near the common thermodynamic equilibrium granted by the H-theorem. The lower-order non trivial equality obtained from the Boltzmann equations leads to a linear functional equation in the velocity variable which is solved thanks to the Fredholm alternative. Since we consider multicomponent mixtures, the classical techniques introduced by Grad cannot be applied, and we propose a new method to treat the terms involving particles with different masses. The next-order equality in the Hilbert expansion then allows to write the macroscopic continuity equations for each component of the mixture.
- In [58], we discuss some numerical properties of the viscous numerical scheme introduced in [23] to solve the one-dimensional pressureless gases system, and study in particular, from a computational viewpoint, its asymptotic behavior when the viscosity parameter $\varepsilon > 0$ used in the scheme becomes smaller.
- In [33] we study a network-based model for rubber. Since the pioneering work by Treloar, many models based on polymer chain statistics have been proposed to describe rubber elasticity. Recently, Alicandro, Cicalese, and the first author rigorously derived a continuum theory of rubber elasticity from a discrete model by variational convergence. The aim of this paper is twofold. First we further physically motivate this model, and complete the analysis by numerical simulations. Second, in order to compare this model to the literature, we present in a common language two other representative types of models, specify their underlying assumptions, check their mathematical properties, and compare them to Treloar's experiments.
- In [63] our aim is to demonstrate the effectiveness of the matched asymptotic expansion method in obtaining a simplified model for the influence of small identical heterogeneities periodically distributed on an internal surface on the overall response of a linearly elastic body. The results of some numerical experiments corroborate the precise identification of the different steps, in particular of the outer/inner regions with their normalized coordinate systems and the scale separation, leading to the model.
- In cancer modeling, to be able to capture the full in-vivo scale, tumors have to be modeled with continuum models. An important step consists in qualitatively and quantitatively comparing agent-based models (which parameters can generally be identified by experiments in vitro) and continuum models. We derived a first 1D continuum model for tumor growth from the cell based model (Drasdo and Hoehme, 2005): it results in a fluid-type model which capture tumor expansion in both diffusive and compact phenotypes. The tumor expands based on the pressure gradient generated by cell proliferation, the latter being hindered by high density or pressure. In [39] this modeled is analyzed mathematically, showing the existence of traveling waves in the different regimes (with or without internal friction and diffusion due to active movement). In particular the incompressible cells limit is very singular and relates to the Hele-Shaw equation. Numerical results confirm the analysis.

SAGE Project-Team

6. New Results

6.1. Parallelism and convergence in iterative linear solvers

6.1.1. Generation of Krylov subspace bases

Participant: Bernard Philippe.

This work was done in collaboration with L. Reichel, from University of Kent, USA (see 8.3.1).

It is published in a journal [19].

Many problems in scientific computing involving a large sparse square matrix A are solved by Krylov subspace methods. This includes methods for the solution of large linear systems of equations with A, for the computation of a few eigenvalues and associated eigenvectors of A, and for the approximation of nonlinear matrix functions of A. When the matrix A is non-Hermitian, the Arnoldi process commonly is used to compute an orthonormal basis for a Krylov subspace associated with A. The Arnoldi process often is implemented with the aid of the modified Gram–Schmidt method. It is well known that the latter constitutes a bottleneck in parallel computing environments, and to some extent also on sequential computers. Several approaches to circumvent orthogonalization by the modified Gram–Schmidt method have been described in the literature, including the generation of Krylov subspace bases with the aid of suitably chosen Chebyshev or Newton polynomials. We review these schemes and describe new ones. Numerical examples are presented.

6.1.2. Parallel Adaptive Deflated GMRES

Participants: Jocelyne Erhel, Bernard Philippe.

This work was done in the context of the joint Inria/ NCSA laboratory on petascale computing (see 8.3.7), and the c2sexa project (see 8.1.3). Computations were done with GENCI supercomputers (see 8.1.6), using the software GPREMS, AGMRES, DGMRES (see 5.7 , 5.8 , 5.9).

It was presented at two conferences [30] [29], is published in proceedings [39] and is submitted (in revision) to a journal [46]. The algorithms are implemented in the software DGMRES and AGMRES, which are freely available in the PETSC repository.

The GMRES iterative method is widely used as Krylov subspace accelerator for solving sparse linear systems when the coefficient matrix is nonsymmetric and indefinite. The Newton basis implementation has been proposed on distributed memory computers as an alternative to the classical approach with the Arnoldi process. The aim of our work here is to introduce a modification based on deflation and augmented techniques. This approach builds an augmented subspace or a preconditioning matrix in an adaptive way to accelerate the convergence of the restarted formulation. It can be combined with preconditioning methods based for example on domain decomposition. In our numerical experiments, we show the benefits of our method to solve large linear systems.

6.1.3. Memory efficient hybrid algebraic solvers for linear systems arising from compressible flows

Participants: Jocelyne Erhel, Bernard Philippe.

This work was done in collaboration with FLUOREM company, in the context of the joint Inria/ NCSA laboratory on petascale computing (see 8.3.7) and the C2S@EXA project (see 8.1.3). Computations were done with GENCI supercomputers (see 8.1.6), using the software GPREMS, AGMRES, DGMRES (see 5.7 , 5.8 , 5.9).

It has been published in a journal [18].

This paper deals with the solution of large and sparse linear systems arising from design optimization in Computational Fluid Dynamics. From the algebraic decomposition of the input matrix, a hybrid robust direct/iterative solver is often defined with a Krylov subspace method as accelerator, a domain decomposition method as preconditioner and a direct method as subdomain solver. The goal of this paper is to reduce the memory requirements and indirectly the computational cost at different steps of this scheme. To this end, we use a grid-point induced block approach for the data storage and the partitioning part, a Krylov subspace method based on the restarted GMRES accelerated by deflation, a preconditioner formulated with the restricted additive Schwarz method and an aerodynamic/turbulent fields split at the subdomain level. Numerical results are presented with industrial test cases to show the benefits of these choices.

6.1.4. Efficient parallel implementation of the fully algebraic multiplicative Aitken-RAS preconditioning technique

Participant: Thomas Dufaud.

This work was done in collaboration with D. Tromeur-Dervout, from ICJ, University of Lyon and has been published in a journal [14].

This paper details the software implementation of the ARAS preconditioning technique [48], in the PETSc framework. Especially, the PETSc implementation of interface operators involved in ARAS and the introduction of a two level of parallelism in PETSc for the RAS are described. The numerical and parallel implementation performances are studied on academic and industrial problems, and compared with the RAS preconditioning. For saving computational time on industrial problems, the Aitken's acceleration operator is approximated from the singular values decomposition technique of the RAS iterate solutions.

6.1.5. An algebraic multilevel preconditioning framework based on information of a Richardson process

Participant: Thomas Dufaud.

This work was done in the context of the C2S@EXA project (see 8.1.3).

It has been presented at a conference [23] and submitted to the proceedings.

A fully algebraic framework for constructing coarse spaces for multilevel preconditioning techniques is proposed. Multilevel techniques are known to be robust for scalar elliptic Partial Differential Equations with standard discretization and to enhance the scalability of domain decomposition method such as RAS preconditioning techniques. An issue is their application to linear system encountered in industrial applications which can be derived from non-elliptic PDEs. Moreover, the building of coarse levels algebraically becomes an issue since the only known information is contained in the operator to inverse. Considering that a coarse space can be seen as a space to represent an approximated solution of a smaller dimension than the leading dimension of the system, it is possible to build a coarse level based on a coarse representation of the solution. Drawing our inspiration from the Aitken-SVD methodology, dedicated to Schwarz methods, we proposed to construct an approximation space by computing the Singular Value Decomposition of a set of iterated solutions of the Richardson process associated to a given preconditioner. This technique does not involve the knowledge of the underlying equations and can be applied to build coarse levels for several preconditioners. Numerical results are provided on both academic and industrial problems, using two-level additive preconditioners built with this methodology.

6.2. Parallel numerical algorithms

6.2.1. High Performance Scientific Computing

Participant: Bernard Philippe.

This work was done in collaboration with several authors, from US, Greece, etc. (see 8.3.1 and 8.2.1).

A book appeared on this subject in 2012 [45] and a chapter of this book is devoted to a historical perspective [44].

This comprehensive text/reference, inspired by the visionary work of Prof. Ahmed H. Sameh, represents the state of the art in parallel numerical algorithms, applications, architectures, and system software. Articles in this collection address solutions to various challenges arising from concurrency, scale, energy efficiency, and programmability. These solutions are discussed in the context of diverse applications, ranging from scientific simulations to large-scale data analysis and mining.
As exascale computing is looming on the horizon while multicore and GPU's are routinely used, we survey the achievements of Ahmed H. Sameh, a pioneer in parallel matrix algorithms. Studying his contributions since the days of Illiac IV as well as the work that he directed and inspired in the building of the Cedar multiprocessor and his recent research, unfolds a useful historical perspective in the field of parallel scientific computing.

6.2.2. Counting eigenvalues in domains of the complex field

Participant: Bernard Philippe.

This work is done in collaboration with E. Kamgnia, from the University of Yaounde 1, Cameroon, in the context of the MOMAPLI project at LIRIMA (see 8.3.5).

It is accepted for publication in a journal [15], and was presented in conferences [31], [32], [38], [40].

A procedure for counting the number of eigenvalues of a matrix in a region surrounded by a closed curve is presented. It is based on the application of the residual theorem. The quadrature is performed by evaluating the principal argument of the logarithm of a function. A strategy is proposed for selecting a path length that insures that the same branch of the logarithm is followed during the integration. Numerical tests are reported for matrices obtained from conventional matrix test sets.

The procedure is now combined with the PPAT methodology (see 5.10). A list of triangles is built for overlapping the boundary of the pseudo-spectra. From the list of vertices, a closed polygonal line is defined and the number of enclosed eigenvalues is determined.

6.2.3. Ratio-Based Parallel Time Integration

Participant: Jocelyne Erhel.

This work is done in in the context of the MODNUM project (see 8.3.2), in collaboration with American University of Beirut (AUB), Lebanon.

It was presented at a conference [41] and is submitted to the proceedings. It was also presented at a seminar of Inria Rennes.

Because time-integration of time-dependent problems is inherently sequential, time parallelism aims mainly at reducing the computational time of some real-time evolutionary problems and may be done through predictor-corrector schemes.

We apply the rescaling method onto initial value problems having an explosive or oscillatory solution, in infinite time. We show how a relevant choice of the end-of-slice condition and the time-rescaling factor might lead to rescaled systems having a uniform convergence to a limit problem. This property provides much better predictions and enhances the relevance of RaPTI that consists mainly of (i) the little sequential computations it involves (predictions and corrections are done in parallel), (ii) the relatively low communication cost it induces and (iii) the similarity of the computation on all slices yielding similar computational times on all processors. Hence, significant speed-ups are achieved. This is illustrated on two problems: a non-linear diffusion-reaction problem having an explosive solution, and a membrane problem having an oscillatory and explosive solution.

6.3. Numerical models and simulations applied to physics

6.3.1. Heat transfer modeling in saturated porous media

Participant: Édouard Canot.

This work is done in the context of the ARPHYMAT project (see 8.3.3) and the MODNUM project (see 8.3.2), in collaboration with Archeosciences, IPR and Lebanese International University (LIU), Lebanon. It was also done in the context of Caroline Thoux's internship (L3, INSA Rennes).

This work is published in [17].

In this paper, the authors introduce a robust numerical strategy to estimate the temperature dependent heat capacity, thermal conductivity and porosity of a saturated porous medium, basing on the knowledge of heating curves at selected points in the medium. In order to solve the inverse problem, we use the least squares criterion (in which the sensitivity coefficients appear), leading to a system of ordinary differential equations (ODE). At the stage of numerical computations, we propose a new global approach, based on the method of lines and ordinary differential equations solvers, combined with a modified Newton method to deal with the nonlinearities presented in the system of coupled equations.

Concerning strong thermal transfer in saturated porous media, the LHA method (Latent Heat Accumulation) is able to take into account phase changes by considering heat accumulation at the local level. The explicit knowledge of the cells which are changing their state allows the build of the liquid-gas interface position. A 2D configuration has been considered, together with a structured mesh but without refinement. The validation of this new method has been checked by making comparison between numerical results and an analytical solution.

6.3.2. Granular materials

Participant: Édouard Canot.

This work is done in collaboration with IPR and is published in [11].

We first studied the granular flows by the "discrete elements" method in silo geometries. By changing the micro-mechanical properties of the grains (restitution and friction) we showed that they had a significant influence on the flow discharge. Although models such as "discrete elements" provide access to all the individual properties of the grains, they have one major drawback: the computation time is very important that prohibits the modeling of geophysical and industrial situations. To overcome this problem, we used the "continuous medium" approach, which consider that the granular medium studied follows a rheology recently proposed in the literature. After discussing the numerical implementation, we have studied this rheology for steady and fully developed flows with a semi-analytical method in two configurations: a shear cell and a channel. This allowed us to highlight the differences between a granular medium and a Newtonian fluid.

6.4. Models and simulations for flow and transport in porous media

6.4.1. Flow and transport in highly heterogeneous porous medium

Participants: Jocelyne Erhel, Grégoire Lecourt, Géraldine Pichot.

This work is done in the context of the H2MNO4 project (see 8.1.1), the H2OGUILDE project (see 8.1.4), the HEMERA project (see 8.1.2). Computations are partly done with GENCI supercomputers (see 8.1.6), using the platform H2OLab (see 5.1) and the software GWNUM, GWUTIL, PARADIS (see 5.3 , 5.2 , 5.5).

This work was done in collaboration with A. Beaudoin, from University of Poitiers (Pprime) and J.-R. de Dreuzy, from Geosciences Rennes (who is on leave until 2013 at UPC, Barcelona, Spain, see 8.2.1). It is also done in collaboration with A. Debussche, from ENS-Cachan-Rennes/Ipso Inria team. It was also done in the context of Grégoire Lecourt's internship (M2, INSA Rennes).

It has been presented at a conference (plenary talk) [26] and a paper is submitted to a journal.

Models of hydrogeology must deal with both heterogeneity and lack of data. We consider in this paper a flow and transport model for an inert solute. The conductivity is a random field following a stationary log normal distribution with an exponential or Gaussian covariance function, with a very small correlation length. The quantities of interest studied here are the expectation of the spatial mean velocity, the equivalent permeability and the macro spreading. In particular, the asymptotic behavior of the plume is characterized, leading to large simulation times and in turn to large physical domains. Uncertainty is dealt with a classical Monte Carlo method, which turns out to be very efficient, thanks to the ergodicity of the conductivity field and to the very large domain. These large scale simulations are achieved by means of high performance computing algorithms and tools.

6.4.2. Solving flow equations in highly heterogeneous porous medium

Participant: Thomas Dufaud.

This work was done in collaboration with L. Berenguer and D. Tromeur-Dervout, from University of Lyon (ICJ).

It is published in a journal [12].

This paper is devoted to the acceleration by Aitken's technique of the convergence of the Schwarz domain decomposition method applied to large scale 3D problems with non separable linear operators. These operators come from the discretization of groundwater flow problems modeled by the linear Darcy equation, where the permeability field is highly heterogeneous and randomly generated. To be computationally efficient, a low-rank approximation of the Aitken's formula is computed from the singular value decomposition of successive iterated solutions on subdomains interfaces. Numerical results explore the efficiency of the solver with respect to the random distribution parameters, and specific implementations of the acceleration are compared for large scale 3D problems. These results confirm the numerical behavior of the methodology obtained on 2D Darcy problems [49].

6.4.3. Transport in discontinuous porous medium

Participants: Lionel Lenôtre, Géraldine Pichot.

This work was done in collaboration with A. Lejay, from Inria Nancy, in the context of the H2MNO4 project (8.1.1).

It is published in a journal [16].

We propose new Monte Carlo techniques for moving a diffusive particle in a discontinuous media. In this framework, we characterize the stochastic process that governs the positions of the particle. The key tool is the reduction of the process to a Skew Brownian Motion (SBM). In a zone where the coefficients are locally constant on each side of the discontinuity, the new position of the particle after a constant time step is sampled from the exact distribution of the SBM process at the considered time. To do so, we propose two different but equivalent algorithms: a two-steps simulation with a stop at the discontinuity and a one-step direct simulation of the SBM dynamic. Some benchmark tests illustrate their effectiveness.

6.4.4. Adaptive stochastic collocation method for an elliptic problem with random data

Participants: Jocelyne Erhel, Mestapha Oumouni.

This work is done in collaboration with Z. Mghazli, from the university of Kenitra, Morocco, in the context of the joint PhD supervision and the HYDRINV project (see 8.3.8 ,8.3.4).

This work has been presented at two conferences [43] [42].

Stochastic collocation methods are frequently used for elliptic equations with random coefficients. However, sparse grid methods are quite expensive and adaptive approaches are designed to save computations.

6.4.5. Reactive transport

Participants: Édouard Canot, Jocelyne Erhel, Souhila Sabit.

This work is done in the context of the MOMAS GNR (8.1.7), the contract with Andra (7.1) and the C2S@EXA project (see 8.1.3). Computations use the software GRT3D (see 5.6).

It has been presented at a conference and a workshop [35] [36].

Modeling reactive transport of contaminants in porous media is a complex time-dependent problem, due to combining the difficulties of modeling transport and chemistry, especially the coupling between them. In this work, we are interested to solve this type of coupling. Several methods have been developed for the resolution for solving this type problem. We choose to solve this problem by a global approach, which considers all the equations as a whole system of differential algebraic equations (DAE), which come from the spatial-only discretization of the equations (method of lines). This approach uses implicit schemes, which imply solving many large linear systems with the Jacobian matrix. The differential algebraic system (DAE) is solved by the solver IDA Sundials. Our new technique is implemented in the GRT3D software; we have observed that the CPU time increases very fast with the size of the system. Our aim is thus to reduce this computation time. Profiling tools have shown that an important part of this computation is due to the linear solving related to the Jacobian matrix. We focus our effort on improving this part, by exploiting the 3x3 block-structure of the Jacobian matrix, via a Gaussian block elimination technique. Our simulations are performed on academic test cases, which involve few chemical components (4 to 5) for both 1D and 2D geometries, giving a number of unknowns up to 72000. First results have shown that our technique is very promising, because the CPU time is reduced by approximately 40 After this part, we eliminated the tracer in our test cases. In GRT3D-SL software, we calculated the concentrations directly without using the Logarithms and with this software, we have reduced the CPU time to 50 %.

6.5. Models and simulations for flow in porous fractured media

This work is done in collaboration with J.-R. de Dreuzy, from Geosciences Rennes (who is on leave until 2013 at UPC, Barcelona, Spain, see 8.2.1). It is done in the context of the GEOFRAC project (see 8.1.5), the H2OGUILDE project (see 8.1.4), the HEMERA project (see 8.1.2), and the Joint Laboratory for Petascale Computing (see 8.3.7). Computations are partly done with GENCI supercomputers (see 8.1.6), using the platform H2OLab (see 5.1) and the software GWNUM, GWUTIL, MPFRAC (see 5.3 , 5.2 , 5.4).

6.5.1. Influence of fracture scale heterogeneity on the flow properties of three-dimensional Discrete Fracture Networks

Participant: Géraldine Pichot.

This work is published in a journal [21].

While permeability scaling of fractured media has been so far studied independently at the fracture- and network- scales, we propose a numerical analysis of the combined effect of fracture-scale heterogeneities and the network-scale topology. The analysis is based on 2×10^6 discrete fracture network (DFNs) simulations performed with highly robust numerical methods. Fracture local apertures are distributed according to a truncated Gaussian law, and exhibit self-affine spatial correlations up to a cutoff scale L_c. Network structures range widely over sparse and dense systems of short, long or widely-distributed fracture sizes and display a large variety of fracture interconnections, flow bottlenecks and dead-ends. At the fracture scale, accounting for aperture heterogeneities leads to a reduction of the equivalent fracture transmissivity of up to a factor of 6 as compared to the parallel plate of identical mean aperture. At the network scale, a significant coupling is observed in most cases between flow heterogeneities at the fracture and at the network scale. The upscaling from the fracture to the network scale modifies the impact of fracture roughness on the measured permeability. This change can be quantified by the measure α_2, which is analogous to the more classical power-averaging exponent used with heterogeneous porous media, and whose magnitude results from the competition of two effects: (i) the permeability is enhanced by the highly transmissive zones within the fractures that can bridge fracture intersections within a fracture plane; (ii) it is reduced by the closed and low transmissive areas that break up connectivity and flow paths.

6.5.2. Synthetic benchmark for modeling flow in 3D fractured media

Participants: Jocelyne Erhel, Géraldine Pichot.

This work is published in a journal [22].

Intensity and localization of flows in fractured media have promoted the development of a large range of different modeling approaches including Discrete Fracture Networks, pipe networks and equivalent continuous media. While benchmarked usually within site studies, we propose an alternative numerical benchmark based on highly-resolved Discrete Fracture Networks (DFNs) and on a stochastic approach. Test cases are built on fractures of different lengths, orientations, aspect ratios and hydraulic apertures, issuing the broad ranges of topological structures and hydraulic properties classically observed. We present 18 DFN cases, with 10 random simulations by case. These 180 DFN structures are provided and fully documented. They display a representative variety of the configurations that challenge the numerical methods at the different stages of discretization, mesh generation and system solving. Using a previously assessed mixed hybrid finite element method (Erhel et al., 2009a), we systematically provide reference flow and head solutions. Because CPU and memory requirements stem mainly from system solving, we study direct and iterative sparse linear solvers. We show that the most cpu-time efficient method is a direct multifrontal method for small systems, while conjugate gradient preconditioned by algebraic multrigrid is more relevant at larger sizes. Available results can be used further as references for building up alternative numerical and physical models in both directions of improving accuracy and efficiency.

6.5.3. Robust numerical methods for solving flow in stochastic fracture networks

Participants: Jocelyne Erhel, Géraldine Pichot.

This work is published in a journal [20] and was presented at a conference (plenary talk) [33].

Working with random domains requires the development of specific and robust numerical methods to be able to solve physical phenomena whatever the generated geometries. Hydrogeology is a typical area of application where one has to face uncertainty about the geometry and the properties of the domain since the available information on the underground media is local, gathered through in-situ experiments with outcrops and wells. From measurements, statistical laws are derived that allow the generation of natural-like random media. The focus of this talk will concern flow in discrete fracture networks. The parameters governing the fractures lengths, shapes, orientations, positions as well as their hydraulic conductivity are stochastic. Our objective is to design robust numerical methods to solve Poiseuille's flow in large and heterogeneous stochastic fracture networks. The first part deals with the meshing strategies required to obtain a good quality mesh for any generated networks. The second part is devoted to numerical techniques to solve the flow equations. A Mortar-like method to deal with nonconforming meshes at the fracture intersections is presented as well as a Schur complement approach to solve the linear system of interest in parallel.

6.5.4. Deflation and Neumann-Neumann Preconditionner for Schur Domain Decomposition Method

Participants: Jocelyne Erhel, Géraldine Pichot.

This work was presented at a conference [34]. A paper is in preparation.

We study a domain decomposition method, which takes advantages from both the direct method and the Preconditioned Conjugate Gradient (PCG). This Schur method reduces the global problem to an interface problem, with a natural domain decomposition based on fractures or fracture packs. We propose an original approach for optimizing the algorithm and a global preconditioning of deflation type. Since the Schur complement S is spd, we apply PCG to solve the linear system $Sx = b$. We use the classical Neumann-Neumann (NN) preconditioner. To gain in efficiency, we use only one Cholesky factorization of the subdomain matrices for the preconditionning and the conjugate gradient steps. We also define a coarse space, based on the subdomain definition, to apply a deflation preconditioner. We do a theoretical complexity study of our algorithm. We use this study, with the numerical data, to compute experimental complexity. We compare the results between several combination for the preconditioner. Then, we confront our results with existing solvers.

6.5.5. Flow in complex 3D geological fractured porous media

Participants: Thomas Dufaud, Jocelyne Erhel, Géraldine Pichot.

This work was presented at a conference [24].

This communication focuses on numerical techniques to compute flow in complex 3D geological fractured porous media, where water can flow both in the rock matrix and in the fractures. This study is an extension of the models designed in the teams SAGE and POMDAPI. The numerical model deals with steady-state flow for single phase and incompressible fluid. In the rock matrix, the flow is governed by Darcy's law, while the flow in the fractures is governed by Poiseuille's law. For both, the law of mass conservation is verified. In a first part, we present the model. Then we propose a test case and its discretization considering a Mixed Hybrid Finite Element Method.

SERPICO Team

6. New Results

6.1. Robust parametric stabilization of moving cells

Participants: Solène Ozeré, Patrick Bouthemy, Charles Kervrann.

Paper under review.

Analysing the dynamic behaviour of individual particles (e.g., proteins, vesicles) inside a cell is of primary importance in cell biology. However, the motion of these particles observed in live cell microscopy image sequences is the addition of the global movement of the cell and their own single motions. Hence, automatically stabilizing the cell (or a group of cells), i.e. compensating for its global motion or equivalently registering its successive positions over time, is previously required. We have proposed a cell stabilization method based on a realtime robust multiresolution scheme (Motion2D software [36]). It can simultaneously handle the estimation of 2D parametric global motions (e.g., affine motion model) and of temporal intensity variations. Three temporal intensity models have been investigated: constant additive model, exponential decay model (corresponding to the photobleaching effect), continuity equation. We have carried out experiments on three biological situations: development of cells, displacements of endosomes, protein recruitment by the Golgi. We have demonstrated the accuracy of our method on these challenging examples and its capacity to efficiently reveal the own motion of subcellular particles. It yields better results than the StackReg method (http://bigwww.epfl.ch/thevenaz/stackreg/), classically used in the field, in cases involving strong local dynamics (see Fig. 6).

Partners:: Perrine Paul-Gilloteaux (UMR 144 CNRS PICT IBiSA Institut Curie)

6.2. Motion classification for interpreting subcellular dynamics

Participants: Antoine Basset, Patrick Bouthemy, Charles Kervrann.

We have just started to address the classification of motions of subcellular particles in light microscopy time-lapse image sequences. For the while, we are considering the following three general classes: diffused motion, obstructed motion and directed motion. We are investigating three approaches. First, we can design likelihood ratio tests for deciding the relevant configuration on local patches. Second, we can define a short-term classification framework based on optical flow computed at time t. The third approach is a mid-term one exploiting pieces of trajectories (tracklets) computed by tracking a set of points.

Partners:: Jérôme Boulanger (UMR 144 CNRS Institut Curie)

6.3. Aggregation of patch-based estimations for illumination-invariant optical flow in live cell imaging

Participants: Denis Fortun, Charles Kervrann, Patrick Bouthemy.

Paper under review.

Live cell image sequences provide a large variety of challenging situations for motion estimation. We have developed a novel optical flow estimation method in the line of work of [11], based on a two-stage aggregation framework and designed to handle this diversity of issues. First, semi-local candidates are estimated with a combination of patch correspondences and illumination-invariant affine motion estimations. Then, one candidate is selected at each pixel in a graph-cut based global aggregation stage. This approach allows us to overcome usual limitations of existing methods such as loss of small structures with large displacements, dependency on illumination changes and over-smoothing of the motion discontinuities. The method jointly estimates the motion and illumination fields. We have compared our approach to state-of-the-art methods and have demonstrated its ability to outperform existing methods in challenging cases frequently arising in live cell imaging (see Fig. 7).

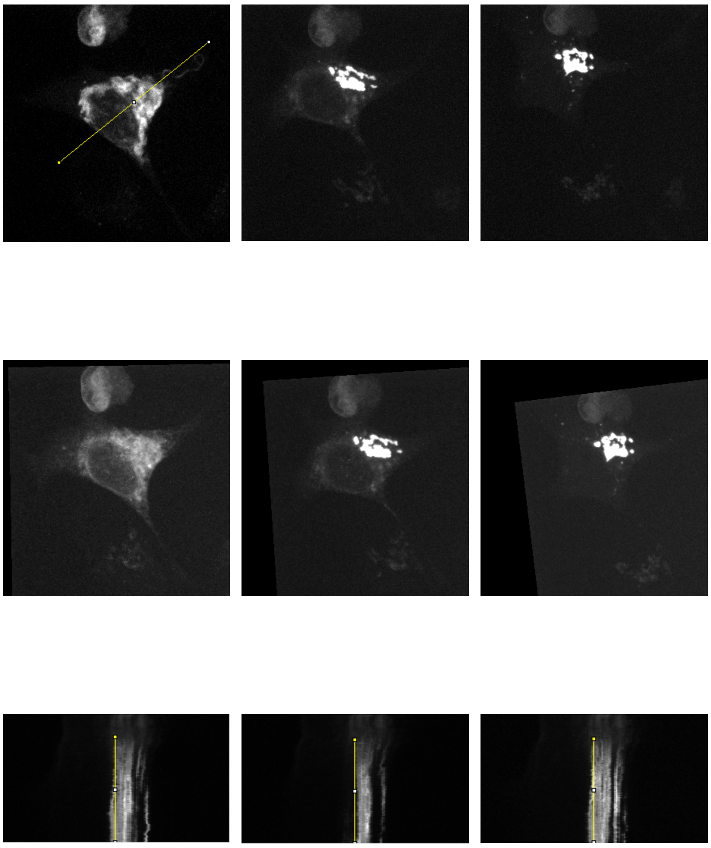


Figure 6. Real-time imaging of the synchronized trafficking of ManII-SBP-EGFP [21]. HeLa cells were transfected to express Ii-streptavidin as a hook and ManII-SBP-EGFP as a reporter. Release of the reporter was induced by addition of biotin and monitored using a spinning disk confocal microscope (F. Pérez, UMR 144 CNRS Institut Curie): 1st row: original images. 2nd row: images out of the motion-compensated sequence at time $t = 20, 60$ and 140 computed with our method; 3rd row: kymographs of the backwarped sequence computed respectively with the baseline motion equation (left), with the exponential decay (middle) and with STACKREG *software (right).*

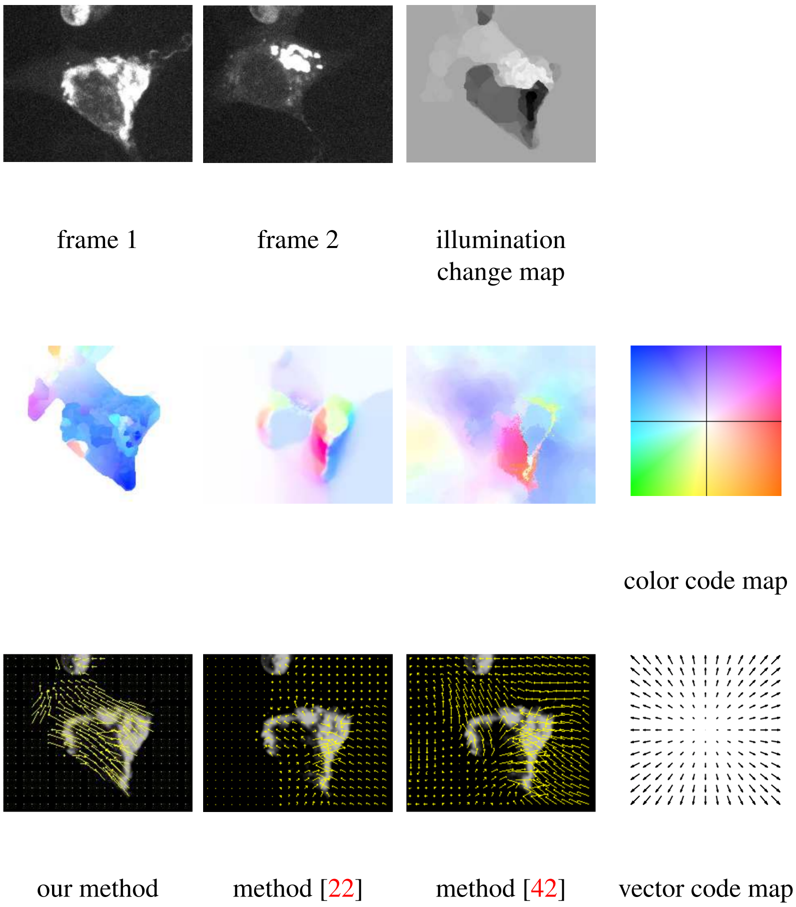


Figure 7. Comparison of our method with the methods of [22] and [42] on a sequence of “HeLa cells” (courtesy of F. Pérez UMR 144 CNRS Institut Curie, PICT-IBiSA).

Partners:: Perrine Paul-Gilloteaux (UMR 144 CNRS PICT IBiSA Institut Curie)

6.4. Correlation and variational approaches for motion and diffusion estimation in fluorescence imaging

Participants: Denis Fortun, Charles Kervrann.

Paper under review.

In this work, we have compared a correlation-based approach and a variational method for both motion and diffusion estimation in representative cell biology studies in fluorescence imaging. The so-called Spatio-Temporal Image Correlation Spectroscopy (STICS) is widely used in fluorescence imaging to recover physical parameters (e.g. direction of flow or Brownian motion of molecules). We have investigated recent advances in variational dense motion estimation and we have proposed to adapt the variational framework to the estimation of diffusion (i.e. Brownian motion). We have demonstrated the influence of the regularization parameter in the variational approach and its ability to capture motion of individual or clusters of moving objects. We have evaluated the advantages and limits of the two approaches for different biological studies (see Fig. 8).

Partners:: Perrine Paul-Gilloteaux, Francois Waharte and Chen Chen (UMR 144 CNRS PICT IBiSA Institut Curie)

6.5. Noise modeling and denoising for intensified camera in fluorescence imaging

Participants: Philippe Roudot, Charles Kervrann.

Two papers under review.

Image intensifiers are commonly used in low light level biological imaging, especially for fluorescence imaging. In this study, we have proposed a statistical framework for noise variance estimation dedicated to image sequences acquired with ICCD (Intensifier CCD). The model has been exploited for fluorescence lifetime estimation (Fluorescence lifetime imaging microscopy, FLIM) [13], [12] and image denoising. We have investigated an alternative approach to [41] and we have shown that intensifier gain variation cannot be neglected in the variance estimation as opposed to a CCD sensor gain. Additionally, we have suggested to correct the noise model spatially to cope with microscopical aberration which are common in experimental setups (see Fig. 9). Finally, we have proposed a novel denoising algorithm based on the NL-means filter [23] which does rely on variance stabilization. The novel patch-based filter is able to adapt to local intensity-based noise statistics (see Fig. 10).

Partners: F. Waharte and J. Boulanger (UMR 144 CNRS PICT IBiSA Institut Curie)

6.6. Microtubules modeling for variational assimilation analysis

Participants: Pierre Allain, Charles Kervrann.

In this project, we propose a bio-physical modeling of growing microtubules at the scale of a single cell. The theoretical advantage of such a modeling is to step aside empirically-based heuristics often carrying artificial parameters which can be hard to tune and to make sense in a data analysis context. We thus propose to model microtubules as rigid and growing cylinders alike (Nedelec and Foethk 2007) [34] but including Newtonian dynamics.

Using both this modeling and fluorescence microscopy, we aim at controlling simulated microtubules to satisfy in vitro observations. We plan to use variational assimilation with adjoint method in the future to achieve such an estimation. We believe that this approach should be able to provide information both on microtubules properties and on vesicle transport dynamics.

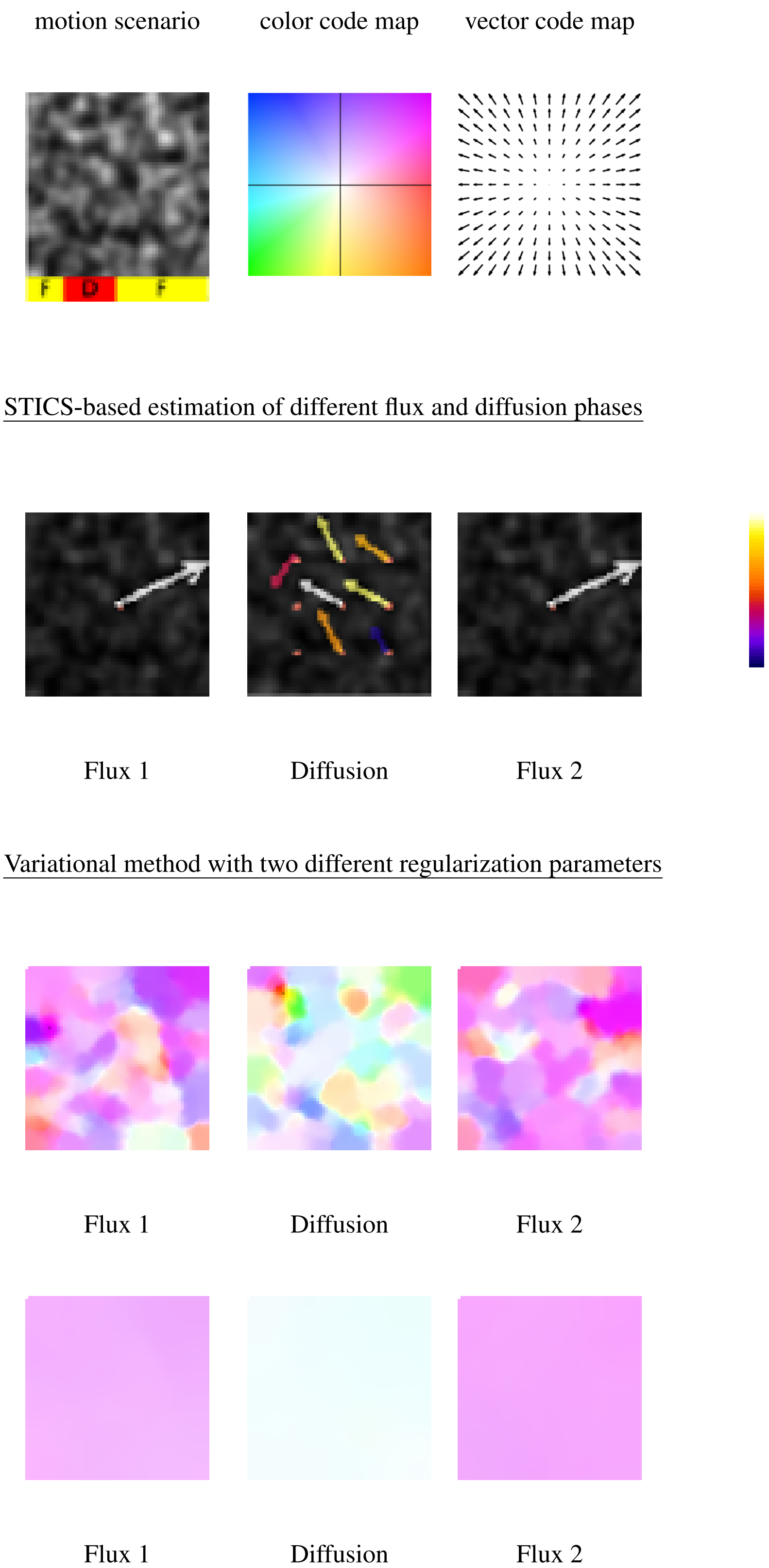


Figure 8. Analysis of STICS and variational methods on artificial image time series with three phases. First row: first frame of the sequence and temporal description of the 3 phases: F/Flux (i.e. directed flow), D/Diffusion (20 - 30 - 40 images) (left); coding maps of vector fields (middle and right). Second row: STICS analysis for each phase. The arrows show the direction of the displacement and the color code is used to represent orientation and magnitude of estimated velocities. Third and fouth rows: Variational estimation for image pairs of each phase with a low regularization parameter (third row) and a high regularization parameter (fourth row).

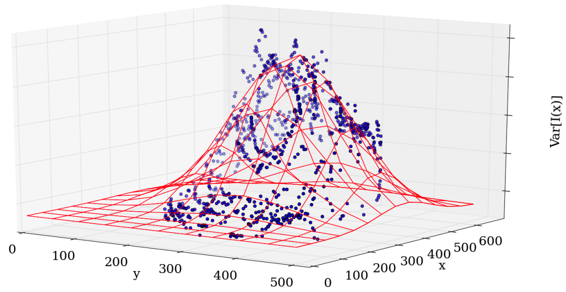


Figure 9. Variance prediction after noise model calibration using a fluorescein FD-FLIM reference stack acquired with a wide-field (WF) microscope (parametric model in red and measurements in blue).

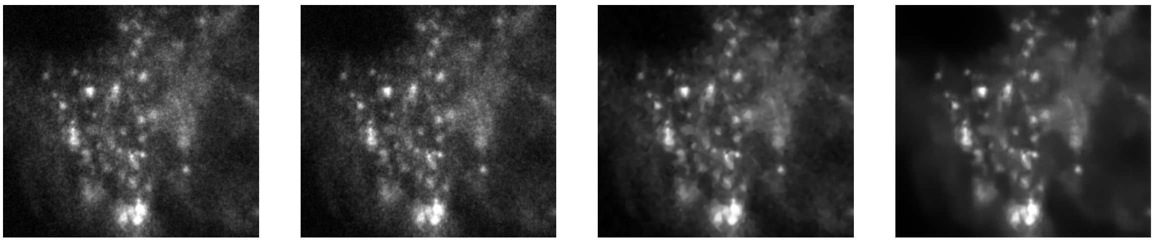


Figure 10. Denoising performance on a live cell image acquired in FD-FLIM (fluorescent tagged caveolin protein) using a confocal setup. From left to right: original image, results with BM3D *[27]* ND-SAFIR *[4] and our method.*

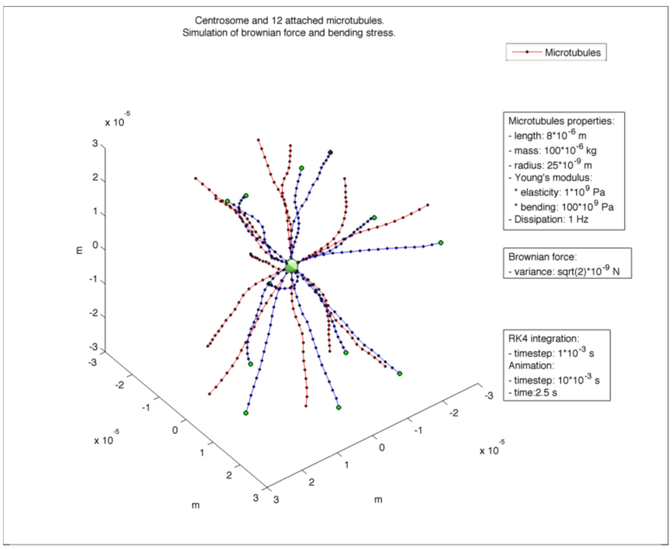


Figure 11. Simulation of growing microtubules in 3D.

6.7. Single versus dual-axis cryo-electron tomography for reconstruction of microtubules assembled in vitro

Participant: Charles Kervrann.

Single-axis cryo-electron tomography of vitrified specimens has become a method of choice to reconstruct in three dimensions macromolecular assemblies in their cellular context or prepared from purified components. In [9], we described a dual-axis acquisition scheme able to improve three-dimensional reconstructions of microtubules assembled in vitro. We showed that in single-axis tomograms, microtubules oriented close to the perpendicular of the tilt axis display diminished contrast, and ultimately transform into sets of parallel lines oriented in the direction of the electron beam when observed in cross-section. We analyzed projections in three-dimensional Fourier transform to demonstrate that imaging artifact is due to a decrease in the angular sampling of their equatorial components. Although the second orthogonal series of images does not fully complement the first one at the specimen level due to increased radiation damage, it still allows elongated features oriented in any directions to be correctly reconstructed, which might be essential for highly heterogeneous specimens such as cells.

Partners: Denis Chrétien, Audrey Guesdon and Sophie Blestel (UMR 6290 CNRS University of Rennes 1)

6.8. Analysis of lateral organization of ordered domains at the plasma membrane surface

Participant: Charles Kervrann.

Paper under review.

In this study, we have analysed a recently designed probe, di-4-ANEPPDHQ, that can change its fluorescent properties depending on whether it is residing in two distinct phases (ordered phase vs disordered phase) of the tobacco cell plasma membranes. We performed a spatial analysis of small (<200 nm) ordered domains observed in multispectral confocal microscopy. We focused on relevant binary images, assumed to be realizations of a MRF-Ising model, depicting the spatial organization of ordered domains. The Ising model depends on 2 parameters: the external field parameter h which controls the total fraction of the "ordered" phase and the interaction parameter which controls the spatial coupling. Maximum pseudo-likelihood methods were investigated to estimate parameters able to describe the spatial properties of ordered domains at the scale of 200 nm $\times$ 200 nm. Almost all estimates of the coupling parameter were positive excluding complete spatial randomness of ordered domains and showing a tendency to spatial aggregation at small distance. We then measured the strength of spatial aggregation through the calculation of the variability fraction explained by the spatial coupling. The mean fraction is low (0.5%) suggesting positive and limited interacting forces between neighbor ordered pixels. Altogether our simulations and analyses provided a probabilistic spatial characterization of PM ordered domains, indicating that recorded images showed a two-scale organization with spatial randomness at large scales (several micrometers) associated with spatial aggregation due to short-range interactions (up to 400 nm).

Partners: P. Gerbeau-Pissot, F. Simon-Plas (UMR 1088 PME INRA, Dijon) and K. Kiêu (MIA Unit INRA, Jouy-en-Josas)

6.9. Line detection in microarray scanner images

Participants: Alice Bergonzoni, Charles Kervrann.

In this study, we have studied two approaches to automatically detect straight lines in images (tool-slide) for calibrating scanners designed by Innopsys company. The Hough transform has been investigated and is able to produce satisfying results provided the algorithm parameters are carefully adjusted (see Fig. 12). To overcome this difficulty, we have evaluated the potential of *a contrario* approach [28] which is well ground theoretically and requires no object prior and parameter adjustment. According to the Helmholtz principle which is based on the *a contrario* approach, any structure is considered in a white noise image as a deviation from randomness.

A meaningful segment is detected when the expectation of its number of occurrences in a white noise image (i.e. number of false alarms) is low. We have evaluated the potential of this method and performed experiments using the LSD algorithm [46] inspired from [28].

Partners:: V. Paveau (Innopsys)

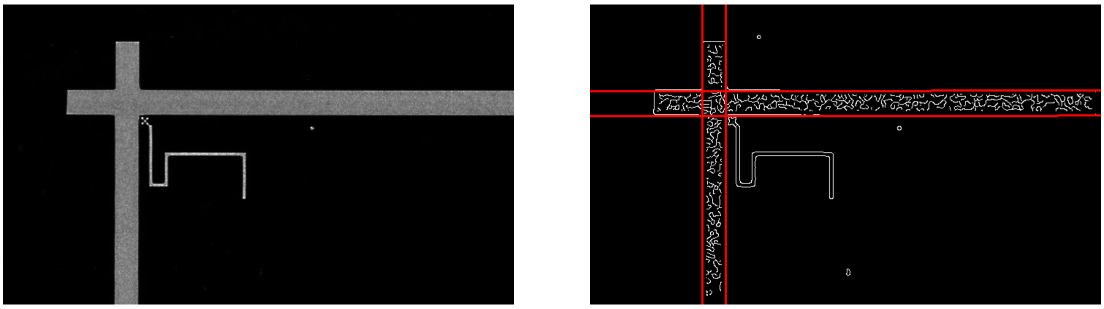


Figure 12. Detection of lines in an image (tool-slide) (pixel size: $3\mu m \times 3\mu m$).

SHACRA Project-Team

6. New Results

6.1. Non-Rigid Augmented Reality for Hepatic Surgery

Hepatic resection and tumors removal approaches remains a major challenge. Despite the use of new minimally invasive techniques which has several advantages such as precision, decreased blood loss, quicker healing time and less pain, the lack of informations due to poor depth perception and direct contact lost leads the surgeons and the research groups to use Augmented Reality to overcome these issues. Augmented Reality is the visual overlay of computers-generated images over real world images. This technique can be used to overlay vessels, tumors and cutting planes performed on the pre-operative data (3D reconstruction from CT or MR scan) onto the laparoscopic video per-operatively. However, current techniques are limited to a rigid registration of the pre-operative liver anatomy onto the intra-operative image, and often this registration is not performed automatically. Our objective is to develop a real-time, non-rigid registration and tracking of the intra and pre-operative liver data.

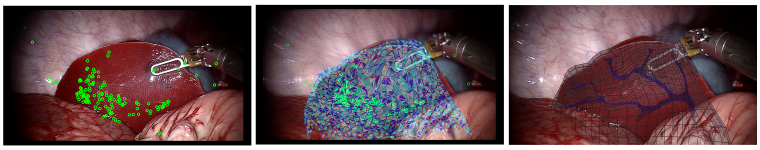


Figure 5. Non-rigid augmentation of a vascular network of a porcine liver : (left) The liver tracking. (Middle) Biomechanical model of the liver under deformation. (Right) Overlaid vascular network.

6.2. Implicit Modeling of Vascular Trees

Many clinical applications require a vessel segmentation process that is able to both extract the centerline and the surface of the blood vessels. However, noise and topology issues (such as kissing vessels) prevent existing algorithms from being able to easily retrieve such a complex system as the brain vasculature. We propose a new blood vessel tracking algorithm that 1) detect the vessel centerline; 2) provide a local radius estimate; and 3) extracts a dense set of points at the blood vessel surface. This algorithm is based on a RANSAC-based robust fitting of successive cylinders along the vessel. Our method was validated against the Multiple Hypothesis Tracking (MHT) algorithm on 10 3DRA patient data of the brain vasculature. Over 30 blood vessels of various sizes were considered for each patient. Our results demonstrated a greater ability of our algorithm to track small, tortuous and touching vessels (96% success rate), compared to MHT (65% success rate). The computed centerline precision was below 1 voxel when compared to MHT. Moreover, our results were obtained with the same set of parameters for all patients and all blood vessels, except for the seed point for each vessel, also necessary for MHT. The proposed algorithm is thereafter able to extract the full intracranial vasculature with little user interaction.

In the context of computer-based simulation, contact management requires an accurate, smooth, but still efficient surface model for the blood vessels. A new implicit model is proposed, consisting of a tree of local implicit surfaces generated by skeletons (*blobby models*). The surface is reconstructed from data points by minimizing an energy, alternating with an original blob selection and subdivision scheme. The reconstructed models are very efficient for simulation and were shown to provide a sub-voxel approximation of the vessel surface on 5 patients.

6.3. Riskmaps in DBS

As discussed in previous sections, Deep Brain Stimulation is a neurosurgical treatment that provides remarkable benefits in neurological movement and affective disorders. It consists in the implantation of a wired electrode deep into the brain. However, the accuracy of the placement is difficult due to brain shifts occuring during the procedure. Due to a potential risk of hemorrhage during the implantation, we specially investigated the brain shift induced motion of the vascular structures. We proposed a method to estimate this motion, based on a physics simulation that consider brain deformation, cerebrospinal fluid and multiple interactions, such as brain-skull contacts etc. The aim is to take it into account during the pre-operative planification step. Thus, we developped a brain-shift aware risk map. It estimate the risk for a trajectory to dissect a vessel. It could help surgeons to choose a safer trajectory for the electrode, and then avoid hemorrhages. The next steps is the use of more complex deformation models.

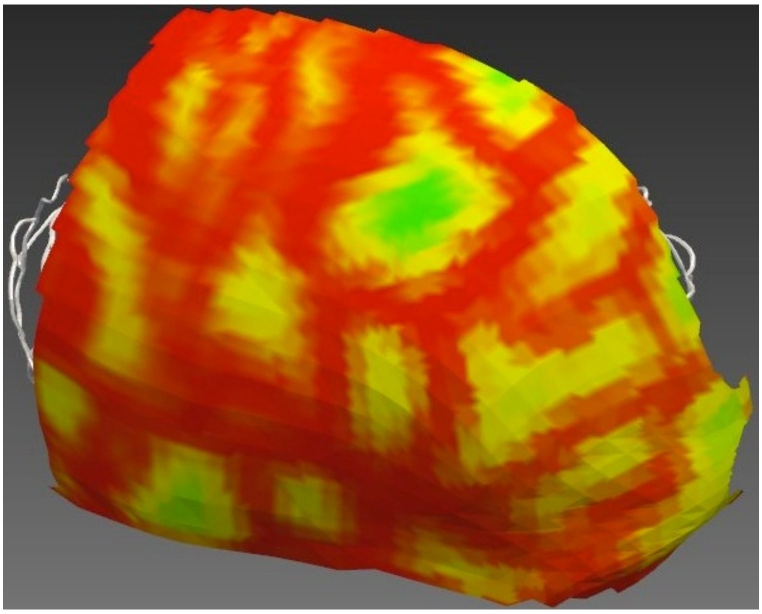


Figure 6. Brain-shift aware risk map

6.4. Electro Physiology

Cardiac arrhythmia is a very frequent pathology that comes from an abnormal electrical activity in the myocardium. This Ph.D. aims at developing a training simulator for interventional radiology and thermo-ablation of these arrhythmias. The latest improvements lead on electrophysiology simulation (using GPU computing) allowed us to reach real-time performance. The issue of fast electrophysiology was a major bottleneck in the development of our simulator.

This new result enabled us to couple the cardiac eletrophysiology with cardiac mechanical models, thus leading to an interactive framework. Our tractable simulation can therefore simulate a patient-specific electrophysiology and then compute the associated cardiac motion using an electromechanical model.

Moreover, the electrophysiology simulation has been also coupled with a navigation simulation. This is still a work in progress. The implementation of more complex models, such as bidomain models, is also in progress.

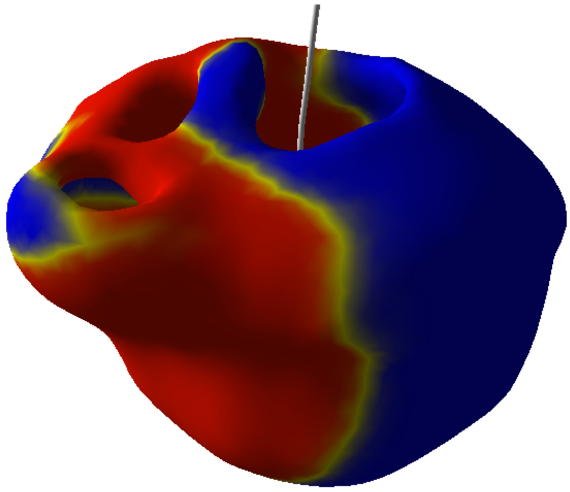


Figure 7. Cardiac electrophysiology computed on a patient-specific geometry

6.5. Shells

Many tissues in human body have thin structure and may be seen as surfaces or at least be modeled as such. Deformation modeling of surfaces is a topic with wide area of applications especially in computer graphics. However, many of the previously presented techniques are not applicable to the area of surgical simulations where a more physically based approach is desired.

To address this problem we present a new model of shell elements based on the formulation of Bézier triangles. To reduce the number of necessary degrees of freedom a kinematic link between nodes inside the element is defined. Furthermore, using implicit integration scheme allows us to achieve interactive frame rate of the simulation.

The applicability of the model has been validated on a prototype of simulator for preoperative planning of surgery of congenital heart diseases.

6.6. Interaction simulation between fluid film and deformable solids

Body fluids are a major constituent of the human body as well by their volume as by their functions. Besides the blood and the lymphatic liquid, many other liquids are present in the body and they have important functions such as lubrication or shock absorption. In this work, we are more particularly interested in the fluids being in the interface between two anatomical structures. We present a method making it possible to simulate the phenomena of interaction between a fluid film and surfaces between which it is forced. The approach that we propose is based on a fluid model and its mechanical coupling with deformable surfaces. According to the pressure of the fluid and the stiffness of the deformable solids in contact with the fluid, various behaviours are expected. Our preliminary results show that it is possible to simulate the main features of these behaviours. Furthermore, the approaches chosen for the fluid model, the deformable model and the coupling between both, are compatible with real time simulations.

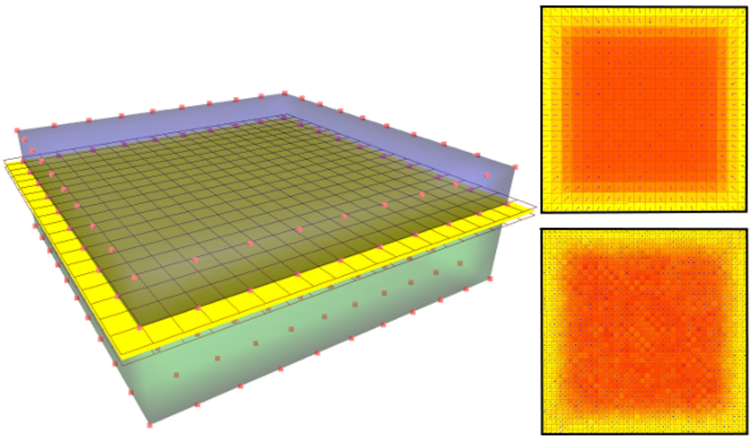


Figure 8. The fluid is between a rigid solid (green) and a deformable solid (blue). The deformable solid is constraint at the edges. Right: the height map of the fluid (yellow minimum and red maximum height).

SISYPHE Project-Team

5. New Results

5.1. Modeling, observation and control: systems modeled by ordinary differential equations

5.1.1. Nonlinear system identification

Participants: Pierre-Alexandre Bliman, Boyi Ni, Michel Sorine, Qinghua Zhang.

In the framework of the joint Franco-Chinese ANR-NSFC EBONSI project, in collaboration with the Laboratory of Industrial Process Monitoring and Optimization of Peking University, and with Centre de Recherche en Automatique de Nancy (CRAN), the topics studied this year on nonlinear system identification are mainly on extended Hammerstein system identification with hysteresis nonlinearity and on continuous time block-oriented nonlinear system identification.

Motivated by the modeling of control valves with significant stiction, we have studied extended Hammerstein systems composed of a hysteresis nonlinearity followed by a linear dynamic subsystem. The joint characterization of the control valve and of the controlled process is formulated as the identification of an extended Hammerstein system. A point-slope based hysteresis model is used to describe the input hysteresis nonlinearity of the control valve. An iterative algorithm is proposed to solve the identification problem. The basic idea is to separate the ascent and descent paths of the input hysteresis nonlinearity subject to oscillatory excitations. Industrial examples are tested to verify the effectiveness of the proposed identification algorithm for characterizing complex behavior of control valve stiction in practice. This work has been presented at the 16th IFAC Symposium on System Identification [66].

A Hammerstein-Wiener system is composed of a dynamic linear subsystem preceded and followed by two static nonlinearities. Typically, the nonlinearities of such a system are caused by actuator and sensor distortions. The identification of such systems with a continuous time model has been studied this year in collaboration with colleagues of CRAN. Based on previously developed simplified refined instrumental variable method, and by making use of an adaptive observer for data filtering, a combined approach, referred to as Kalman pre-filtered instrumental variable based method, is developed. By taking advantages of the two aforementioned methods, the new method is faster and has a naturally stabilizing Kalman filter that does not color white noises. This work has been presented at the 16th IFAC Symposium on System Identification [62].

5.1.2. Model-based fault diagnosis

Participants: Abdouramane Moussa Ali, Qinghua Zhang.

The increasing requirements for higher performance, efficiency, reliability and safety of modern engineering systems call for continuous research investigations in the field of fault detection and isolation. This year we have studied algebro-differential systems through an adaptive observer based approach, and linear time varying systems through a Kalman filter based statistical testing approach.

In the framework of the MODIPRO project funded by Paris Region ASTech, the monitoring of the air conditioning system of an aircraft has been studied this year. Part of this system is modeled by nonlinear algebro-differential equations. A method for fault diagnosis of such systems has been developed in our study. Through a particular discretization method and under realistic assumptions, the considered continuous time DAE model is transformed to an explicit state space model in discrete time. An adaptive observer is then applied to the discretized system for monitoring faults possibly affecting the system and represented by changes in model parameters. This work will be presented at the 5th IFAC Symposium on System Structure and Control [61].

While the theory of fault diagnosis has been mostly developed for linear time invariant (LTI) systems, in many industrial applications it is important to take into account the nonlinear behavior of the monitored systems. One possible approach is to linearize a nonlinear system all along its state trajectory, resulting in linear time varying (LTV) or linear parameter varying (LPV) models. In collaboration with Michèle Basseville of IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires), fault diagnosis for stochastic LTV systems has been studied this year. By applying the Kalman filter in a particular manner avoiding the difficulty related to unknown faults possibly affecting the system, the problem of fault diagnosis in a dynamic LTV system is transformed into a hypothesis testing problem in a simple linear regression model. Generalized likelihood ratio (GLR) tests are then applied to the resulting hypothesis testing problem. This work has been presented at the 16th IFAC Symposium on System Identification [67].

5.2. Observation, control and traveling waves in systems modeled by partial differential equations

5.2.1. Modeling of electric transmission networks

Participants: Mohamed Oumri, Michel Sorine, Qinghua Zhang.

The increasing number and complexity of wired electric networks in modern engineering systems is amplifying the importance of the reliability of electric connections. In the framework of the ANR 0-DEFECT project, we have studied mathematical models of complex electric networks with the aim of designing algorithms for fault diagnosis. The well known Baum-Liu-Tesche (BLT) equation is a powerful model for describing quite general networks and allows to compute the current and voltage waves at the nodes of a network from the specifications of its nodes and connecting cables [63]. This year we have studied the inverse problem: what can we know about the properties of the cables connecting the nodes of a network from experiments made at the nodes of the network? A convenient model for this purpose is formulated with admittance matrices. It is essentially equivalent to the BLT equation, hence can describe quite general networks. The inverse problem is then solved through a decomposition of the admittance matrix of the entire network.

5.2.2. Diagnosis of insulator degradation in long electric cables

Participants: Leila Djaziri, Michel Sorine, Qinghua Zhang.

For the diagnosis of insulator degradation in long electric cables, the estimation of the shunt conductance of such cables have been studied, in the framework of the ANR INSCAN project. The shunt conductance of a healthy electric cable is usually very weak. Even when the insulator in the cable is significantly degraded, the shunt conductance can still remain at a quite low level. The main difficulty in this study is due to the fact that the measurements made at the ends of a cable are hardly sensitive to the variations of the shunt conductance. To overcome this difficulty, two methods have been studied. One of them is based on the analysis of the sensitivity of the wave phase shift to the shunt conductance. The efficiency of this method has been demonstrated through extensive tests on cables of SNCF (Société Nationale des Chemins de Fer français). Another method is based on the processing of long time data records. It is designed for the estimation of distributed shunt conductance, in order to detect and to locate inhomogeneous degradation of the insulator. The main idea of this method is to compensate the weak sensitivity of the measurement by long time data records. The results of this method evaluated by numerical simulations have been reported at the 16th IFAC Symposium on System Identification [68].

5.3. System theory approach of some quantum systems

Participants: Hadis Amini, Zaki Leghtas, Mazyar Mirrahimi, Pierre Rouchon.

Most of this work is done in close collaboration with the Pierre Aigrain laboratory (LPA) at ENS Paris and the Quantronics Laboratory (Qlab) of Michel Devoret and the Rob Schoelkopf Lab at Yale University.

Modern scientific and technologic requirements have led the theoretical and experimental research toward an engineering of quantum systems. The technologies that are proposed or developed include nano-scale electromechanical devices, tools for implementing quantum computation and quantum communication, NMR applications, quantum chemistry synthesis, high-resolution sensors, etc. The recent theoretical and experimental researches have shown that the quantum dynamics can be studied in the framework of the theory of estimation and control of systems, but give place to models that are not completely explored yet.

Our activities lie in the theoretical and experimental interface of this progressing field of research with an accent on the applications in quantum information and computation as well as high-precision metrology. By focusing on two different but similar types of experimental setups, consisting of cavity quantum electrodynamical systems and quantum Josephson circuits, we aim in preparing highly non-classical states of a microwave field and protect these states against decoherence. Two different approaches are considered: 1- real-time measurement, quantum filtering and feedback ; 2- dissipation engineering also called reservoir engineering. Through the first methodology, we try to propose new experimental feedback protocols based on a fast real-time processing of measurement signal, followed by a state estimation applying the filtered signal and finally designing simple feedback laws based on the estimated state. The second methodology consists of designing new quantum circuit schemes that allow to orient the system's coupling to its environment in such a way that evacuates the undesired entropy induced by un-controlled noise sources.

5.3.1. Measurement based feedback

We have developed the mathematical methods underlying a recent quantum feedback experiment stabilizing photon-number states [17], [30], [29], [24]. We consider a controlled system whose quantum state, a finite dimensional density operator, is governed by a discrete-time nonlinear Markov process. In open-loop, the measurements are assumed to be quantum non-demolition (QND) measurements. This Markov process admits a set of stationary pure states associated to an orthonormal basis. These stationary states provide martingales crucial to prove the open-loop stability: under simple assumptions, almost all trajectories converge to one of these stationary states; the probability to converge to a stationary state is given by its overlap with the initial quantum state. From these open-loop martingales, we construct a supermartingale whose parameters are given by inverting a Metzler matrix characterizing the impact of the control input on the Kraus operators defining the Markov process. This supermartingale measures the "distance" between the current quantum state and the goal state chosen from one of the open-loop stationary pure states. At each step, the control input minimizes the conditional expectation of this distance. It is proven that the resulting feedback scheme stabilizes almost surely towards the goal state whatever the initial quantum state. This state feedback takes into account a known constant delay of arbitrary length in the control loop. This control strategy is proved to remain also convergent when the state is replaced by its estimate based on a quantum filter. It relies on measurements that can be corrupted by random errors with conditional probabilities described by a known left stochastic matrix. Closed-loop simulations corroborated by experimental data illustrate the interest of such nonlinear feedback scheme for the photon box [29].

We have also investigated the stabilization of the dynamical state of a superconducting qubit. In a series of papers, A. Korotkov and his co-workers suggested that continuous weak measurement of the state of a qubit and applying an appropriate feedback on the amplitude of a Rabi drive, should maintain the coherence of the Rabi oscillations for arbitrary time. Here, in the aim of addressing a metrological application of these persistent Rabi oscillations, we explore a new variant of such strategies. This variant is based on performing strong measurements in a discrete manner and using the measurement record to correct the phase of the Rabi oscillations. Noting that such persistent Rabi oscillations can be viewed as an amplitude- to-frequency convertor (converting the amplitude of the Rabi microwave drive to a precise frequency), we propose another feedback layer consisting of a simple analog phase locked loop to compensate the low frequency deviations in the amplitude of the Rabi drive [60].

5.3.2. Dissipation engineering

We have introduced a new quantum gate that transfers an arbitrary state of a qubit into a superposition of two quasi-orthogonal coherent states of a cavity mode, with opposite phases. This qcMAP gate is based on conditional qubit and cavity operations exploiting the energy level dispersive shifts, in the regime where they are much stronger than the cavity and qubit linewidths [77], [26]. The generation of multi-component superpositions of quasi-orthogonal coherent states, non-local entangled states of two resonators and multi-qubit GHZ states can be efficiently achieved by this gate. We also propose a new method, based on the application of this gate, to autonomously correct for errors of a logical qubit induced by energy relaxation. This scheme encodes the logical qubit as a multi-component superposition of coherent states in a harmonic oscillator. The error correction is performed by transferring the entropy to an ancila qubit and reseting the qubit. We layout in detail how to implement these operations in a practical system [78]. This proposal directly addresses the task of building a hardware-efficient and technically realizable quantum memory [78].

We have also studied the application of dissipation engineering techniques to perform a high-performance and fast qubit reset. Qubit rest is crucial at the start of and during quantum information algorithms. Our protocol, nicknamed DDROP (Double Drive Reset of Population) is experimentally tested on a superconducting transmon qubit and achieves a ground state preparation of at least 99.5$\%$ in times less than 3μs; faster and higher fidelity are predicted upon parameter optimization [74]. We are currently working on extending our protocol to prepare and protect two-qubit entangled states and to perform autonomous quantum error correction.

5.4. Modeling, observation and control in biosciences - Reproductive system

5.4.1. Numerical simulation of the selection process of the ovarian follicles

Participants: Benjamin Aymard, Frédérique Clément.

Collaboration with Frédéric Coquel and Marie Postel.

Implementation of a parallelized numerical scheme based on finite volumes. We have designed and implemented a numerical method to simulate a multiscale model describing the selection process in ovarian follicles [11], [10]. The PDE model consists in a quasi-linear hyperbolic system of large size, namely $N_f \times N_f$, ruling the time evolution of the cell density functions of N_f follicles (in practice N_f is of the order of a few to twenty). These equations are weakly coupled through the sum of the first order moments of the density functions. The time-dependent equations make use of two structuring variables, age and maturity, which play the roles of space variables. The problem is naturally set over a compact domain of $\mathbf{R}^2$. The formulation of the time-dependent controlled transport coefficients accounts for available biological knowledge on follicular cell kinetics. We introduce a dedicated numerical scheme that is amenable to parallelization, by taking advantage of the weak coupling. Numerical illustrations assess the relevance of the proposed method both in term of accuracy and HPC achievements [32].

A numerical method for cell dynamics: kinetic equations with discontinuous coefficients. The motivation of this work is the numerical treatment of the mitosis in biological models involving cell dynamics. More generally we study hyperbolic PDEs with flux transmission conditions at interfaces between subdomains where coefficients are discontinuous. A dedicated finite volume scheme with a limited high order enhancement is adapted to treat the discontinuities arising at interfaces. The validation of the method is done on 1D and 2D toy problems for which exact solutions are available, allowing us to do a thorough convergence study. A simulation on the original biological model illustrates the full potentialities of the scheme [72].

5.4.2. Optimal control of cell mass and maturity in a model of follicular ovulation

Participants: Frédérique Clément, Peipei Shang.

Collaboration with Jean-Michel Coron

We have studied some optimal control problems associated with a scalar hyperbolic conservation law modeling the development of ovarian follicles. Changes in the age and maturity of follicular cells are described by a 2D conservation law, where the control terms act on the velocities. The control problem consists in optimizing the follicular cell resources so that the follicular maturity reaches a maximal value in fixed time. Formulating the optimal control problem within a hybrid framework, we have proved necessary optimality conditions in the form of Hybrid Maximum Principle [36]. We have then derived the optimal strategy and shown that there exists at least one optimal bang-bang control with one single switching time.

5.4.3. Multiscale analysis of mixed-mode oscillations in a phantom bursting model

Participants: Frédérique Clément, Mathieu Desroches, Maciej Krupa, Alexandre Vidal.

We have studied mixed mode oscillations in a model of secretion of GnRH (gonadotropin releasing hormone). The model is a phantom burster consisting of two feedforward coupled FitzHugh-Nagumo systems, with three time scales. The forcing system (Regulator) evolves on the slowest scale and acts by moving the slow null-cline of the forced system (Secretor). There are three modes of dynamics: pulsatility (transient relaxation oscillation), surge (quasi steady state) and small oscillations related to the passage of the slow null-cline through a fold point of the fast null-cline. We have derived a variety of reductions, taking advantage of the mentioned features of the system. We have obtained two results; one on the local dynamics near the fold in the parameter regime corresponding to the presence of small oscillations and the other on the global dynamics, more specifically on the existence of an attracting limit cycle. Our local result is a rigorous characterization of small canards and sectors of rotation in the case of folded node with an additional time scale, a feature allowing for a clear geometric argument. The global result gives the existence of an attracting unique limit cycle, which, in some parameter regimes, remains attracting and unique even during passages through a canard explosion [43].

5.4.4. A network model of the periodic synchronization process in the dynamics of calcium concentration in GnRH neurons

Participants: Frédérique Clément, Maciej Krupa, Alexandre Vidal.

Mathematical neuroendocrinology is a branch of mathematical neurosciences that is specifically interested in endocrine neurons, which have the uncommon ability of secreting neurohormones into the blood. One of the most striking features of neuroendocrine networks is their ability to exhibit very slow rhythms of neurosecretion, on the order of one or several hours. A prototypical instance is that of the pulsatile secretion pattern of GnRH (gonadotropin releasing hormone), the master hormone controlling the reproductive function, whose origin remains a puzzle issue since its discovery in the seventies. We have investigated the question of GnRH neuron synchronization on a mesoscopic scale and study how synchronized events in calcium dynamics can arise from the average electric activity of individual neurons. We have used as reference seminal experiments performed on embryonic GnRH neurons from rhesus monkeys, where calcium imaging series were recorded simultaneously in tens of neurons, and which have clearly shown the occurrence of synchronized calcium peaks associated with GnRH pulses, superposed on asynchronous, yet oscillatory individual background dynamics [100]. We have designed a network model by coupling 3D individual dynamics of FitzHugh-Nagumo type. Using phase-plane analysis, we have constrained the model behavior so that it meets qualitative and quantitative specifications derived from the experiments, including the precise control of the frequency of the synchronization episodes. In particular, we have shown how the time scales of the model can be tuned to fit the individual and synchronized time scales of the experiments. Finally, we have illustrated the ability of the model to reproduce additional experimental observations, such as partial recruitment of cells within the synchronization process or the occurrence of doublets of synchronization [76].

5.5. Clinical and physiological applications

5.5.1. DynPeak: An algorithm for pulse detection and frequency analysis in hormonal time series

Participants: Frédérique Clément, Claire Médigue, Alexandre Vidal, Qinghua Zhang.

Collaboration with Stéphane Fabre (UMR CNRS-INRA 6175).

The endocrine control of the reproductive function is often studied from the analysis of luteinizing hormone (LH) pulsatile secretion by the pituitary gland. Whereas measurements in the cavernous sinus cumulate anatomical and technical difficulties, LH levels can be easily assessed from jugular blood. However, plasma levels result from a convolution process due to clearance effects when LH enters the general circulation. Simultaneous measurements comparing LH levels in the cavernous sinus and jugular blood have revealed clear differences in the pulse shape, the amplitude and the baseline. Besides, experimental sampling occurs at a relatively low frequency (typically every 10 min) with respect to LH highest frequency release (one pulse per hour) and the resulting LH measurements are noised by both experimental and assay errors. As a result, the pattern of plasma LH may be not so clearly pulsatile. Yet, reliable information on the InterPulse Intervals (IPI) is a prerequisite to study precisely the steroid feedback exerted on the pituitary level. Hence, there is a real need for robust IPI detection algorithms. We have designed an algorithm for the monitoring of LH pulse frequency, basing ourselves both on the available endocrinological knowledge on LH pulse (shape and duration with respect to the frequency regime) and synthetic LH data generated by a simple model [54]. We make use of synthetic data to make clear some basic notions underlying our algorithmic choices. We focus on explaining how the process of sampling affects drastically the original pattern of secretion, and especially the amplitude of the detectable pulses. We then describe the algorithm in details and perform it on different sets of both synthetic and experimental LH time series. We further comment on how to diagnose possible outliers from the series of IPIs which is the main output of the algorithm.

STEEP Exploratory Action

6. New Results

6.1. Calibration of TRANUS Adjustment Parameters

One of the most difficult steps in calibrating the parameters of the TRANUS land use model, concerns the estimation of its adjustment parameters (so-called shadow prices), that allow to "absorb" imperfections of the model or the data. The main difficulties are the non-linearity of the underlying equations and the fact that some of these equations give rise to loops between intermediate system variables: modifications of some of these variables entail modifications of others and vice-versa. In other words, the concerned part of TRANUS is a dynamic system. Currently, users of TRANUS perform the calibration by semi-automatic (at best) trial-and-error.

We have started investigating more systematic solutions to this. A first step has been to explicitly pose the estimation problem in the form of an optimization problem, with clearly stated cost function and constraints. Next, we have found ways of splitting the problem into separable subproblems, concerning the estimation of adjustment parameters for different economic sectors. In particular, the housing/land sectors can be calibrated independently of the others. A simple gradient descent was shown to be sufficient, both theoretically and experimentally, to achieve this calibration. We are currently investigating strategies to estimate the adjustment parameters of the remaining sectors.

6.2. Calibration of TRANUS Using Maximum Likelihood Estimation

Calibration of the TRANUS land use module typically involves determination of key parameters which dictate land use assignments and prices. As mentioned earlier, It is a difficult task to calibrate a LUTI model as the number of parameters involved are large and are uncertain. Traditionally, these models are calibrated manually by experts, who try to estimate the parameters using their prior experience. However, such a method is difficult as well as time consuming, especially when the parameter space is large and uncertain. Hence, an algorithmic procedure to estimate parameters from mathematical model is desired.

We have proposed an algorithm to calibrate the land use module of TRANUS using maximum likelihood estimation (MLE). The observed outputs of the land use module is modeled to follow a Gaussian process. The covariance matrix is represented as a function of inputs of the land use module and hyperparameters. A MLE optimization problem is then formulated to estimate the parameters of the land use module and the hyperparameters of the Gaussian covariance kernel. The resulting nonlinear programming (NLP) problem is then solved using NLP solvers based on sequential quadratic programming.

The proposed calibration algorithm has been successfully applied to the model of Grenoble, France ; and the performance of the proposed calibration methodology, has been compared to traditional calibration techniques. The metric to judge performance is assumed to be the $\mathcal{L}_2$ norm of the difference between observed and calculated land use assignments obtained using the calibrated model.

Before this calibration task is performed, a sensitivity analysis has been carried out. Hence, sensitivity analysis of the parameters on the output is important as is helps us identify major sources of uncertainty in terms of their contribution towards output space variability. Here, the *total effect* of the land use parameters on a *quantity of interest* or *QoI* is assessed. The *QoI* is assumed to be the $\mathcal{L}_2$ norm of the difference between observed and calculated land use assignments. For this Grenoble model, the number of uncertain parameters involved are 100, and finally is is observed that only 3 amongst them contribute towards 99.2% of *QoI* variability. [14], [13]

6.3. Material flows, production and consumption at sub-national geographic levels

As explained earlier, estimating the actual environmental impact of an urban area on the one hand, and the efficiency of (local or national) policy options in reducing these impacts on the other, requires an understanding of the material flows and material uses generated by the considered urban area. It is important to realize that impacts (both local and distant) can vary greatly from one region or departement to the next, depending on its agricultural and industrial characteristics. The whole point of this work is to evaluate as best as possible these variations, in order to best adapt public policies in terms of environmental impacts, for given socio-economic conditions and objectives.

The first step in this analysis is to establish a database of production, consumption and exchanges (import and export) at the various geographic levels of interest, and for the various types of material of interest. In practice, the finest scale of available data is a French *département*, and the publicly available data refer to the national, regional or "departemental" level. Only major primary materials are accounted for, through the content of end products and waste in these primary materials (toxic waste are accounted for separately). For example, for cereals such as wheat, production at the departement level is available through the national *Agreste* database, variations of stock are small once averaged over a few years period, import and export are obtained from the *Sitram* database (a database initially elaborated by the ministry of transportation and now maintained by the Ministry of Environment), which follows all national and international transport by transportation mode and by type, through annual stratified polls of transportation companies. Productions of non-agricultural products in France is very low except for construction materials (most notably cement), for which the industry maintains its own publicly available database. Following transformations requires information from various industrial sectors, e.g., the flour trade and food industry for wheat use, taking into account animal farming which consumes a non-negligible fraction of primary agricultural products.

Once this database is constructed, one also needs to estimate production, consumption and imports and exports at finer scale than the departement. In practice, this is performed by correlating the desired information at the national, regional and departemental scale with another auxiliary quantity serving as proxy, that is also known at the desired smaller scale. For example, wheat production can easily be correlated with available surfaces in wheat growing areas, that are known from the Corine Land Cover database at scales of the order of a few hundred meters. More generally, auxiliary quantities are constructed from relevant demographic and economic and geographic data, that are mostly available through the various INSEE databases. This requires some educated guess-work to find the most likely auxiliary quantities, and evaluate their correlation with the quantities of interest at scales where data on both are available. This aspect of the problem has been completed only for food staples at this stage.

An important aspect of the problem is to estimate the errors in the data. Errors can be detected when quantities of a given material are not conserved through transportation and transformation processes. It appears that the largest source of error comes from the transportation database, because the stratified polling methodology is optimized with respect to total transport from a pair of origin and destination, independently of the nature of the transported goods. It is in principle possible to compute confidence intervals per type of material and not only on total volumes of exchanges, but this requires access to some non public information. Discussions have been initiated with the Ministry to have access to this information, in order to estimate the reliability of this method of transport quantification. If the *Sitram* database turns out to be too imprecise, the method described above to estimate lacking data can be applied to transport as well with appropriate auxiliary quantities, but the results also suffer from various sources of error.

This first stage of the Material Flow analysis is nevertheless largely underway. The two next steps consist in environmental impact evaluation on the one hand, at the present date, and in developing a method of analysis of changes of such impacts under various policy scenarios and options. Bith will rely on the use of Life Cycle Analysis databases, as mentioned above.

6.4. Computer vision

Three of our permanent staff have previously been active in computer vision. This activity is gradually coming to an end: the last PhD student has defended his thesis in 2012 and no new projects are started. Since this topic is not central to STEEP, results are only summarized very briefly. The main scientific result has been the development of a novel approach for 3D modeling of semi-transparent objects, which couples both, geometric and photometric information [1]. This allows 3D modeling with fewer input images than previously and potentially, with a higher accuracy. Besides this, our main activity in computer vision has been related to industrial projects, the main goal being to finalize our work of the last years with an industrial transfer.

VIRTUAL PLANTS Project-Team

5. New Results

5.1. Analysis of structures resulting from meristem activity

5.1.1. Acquisition and design of plant geometry

Participants: Chakkrit Preuksakarn, Mathilde Balduzzi, Frédéric Boudon, Christophe Pradal, Christophe Godin, Christian Fournier.

This research theme is supported by RTRA project named PlantScan3D.

Virtual 3D model of plants are required in many areas of plant modeling. They can be used for instance to simulate physical interaction of real plant structures with their environment (light, rain, wind, pests, ...), to set up initial conditions of growth models or to assess their output against real data. In the past decade, methods have been developed to digitize plant architectures in 3D [48], [42]. These methods are based on direct measurements of position and shape of every plant organ in space. Although they provide accurate results, they are particularly time consuming. More rapid and automated methods are now required in order to collect plant architecture data of various types and sizes in a systematic way. In this aim, we explore the use of pictures, laser scanner, video and direct sketching.

- *Automated reconstruction of plant architecture* (Chakkrit Preuksakarn, Mathilde Balduzzi, Julien Diener, Frédéric Boudon, Jean-Baptiste Durand, Christophe Godin, Bernard Mourrain [Inria, Galaad], Franck Hetroy [Inria, Morpheus], Marie-Paule Cani [Inria, Imagine], Pascal Ferraro [Labri, Bordeaux])

  We investigate the possibility to use 3D laser scanners to automate plant digitizing. We are developping algorithms to reconstruct branching systems without leaves or foliage from scanner data or from scan simulated on plant mock-up obtained using different digitizing method. For this we collaborate with the EPI Galaad from Sophia-Antipolis, the EPI Imagine from Grenoble, different INRA teams, UMR PIAF in Clermont Ferrand, UMR LEPSE and AFEF team in Montpellier and Lusignan, the University of Helsinki, Finland and the CFCC in England. We developed a reconstruction pipeline composed of several procedures. A contraction procedure, first aggregates points at the center of the point cloud. The team proposed a simple adaptive scheme to contract points. In a second step, a skeleton procedure uses a Space Colonization Algorithm [47] to build the skeleton of the shape from the contracted point set. This method is adaptive to the local density of the point set. Then a pipe-model based procedure makes it possible to estimate locally diameters of the branches. Finally, an evaluation procedure has been designed to assess the accuracy of the reconstruction and a comparison with alternative methods has been carried out. Publication of this work is in progress.

  An automated reconstruction pipeline is also developed for processing 2D images of root system architecture (RSA) in the context of the Rhizopolis project. The analysis of these data is currently a major challenge in understanding root development. Existing tools either focus on specific applications, on simple structures (for example one root segment) or require long manual work. Here, we develop a processing pipeline that takes as an input 2D high resolution images of petri plates containing root systems. The pipeline makes it possible to extract from the images the whole architecture of root systems, with minimal or no user intervention. In order to obtain this result, the problem was decomposed in several steps: filter and label the input image, extract the image skeleton as a general graph structure and then convert it into a tree structure representing the visualised RSA, using a priori knowledge to solve inconsistencies. The pipeline has been added to the OpenAlea platform, thus allowing resulting data to be directly processed by other advanced high-level computational or statistical tools. The developed pipeline is currently being tuned and tested on several databases of 2D images with varying complexities of both arabidopsis and rice.

Figure 3. Reconstruction of a cherry tree. Left: photograph of the original tree. Right: 3D reconstruction from a laser scan rendered and integrated on the same background.

Additionnally, we also investigate the reconstruction of tree foliage from laser scanners in the context of Mathild Balduzzi's PhD thesis. Such elements are crucial to study the interaction of the plant with its environment. However, laser scans contain outliers on the silhouette of the scans that make the meshing of the point set difficult. New generation of laser scanners provides intensity of the laser reflected) on the surface of scanned objects. This intensity is dependant of the distance to the object, its optical property and the incidence angle. A first work on this topic shows that after correcting the distance effect, the incidence angle can be deduced from the intensity. From this result, we develop a reconstruction technique using the scan intensities and based on Shape-From-Shading approaches. The idea is to generate a new point set from the intensities and a set of seed points. This new point set has the property of being smooth but is not necessarly the exact representation of the scanned object. To consolidate the reconstruction, we are working on merging it with the orignal noisy point set coming from the scans using Kalman filtering. As a result, a final point set will be obtained without noise and with outiers naturally removed.

- *Sketching of plants.* (Frédéric Boudon, Christophe Godin, Steven Longuay [University of Calgary, Canada], Przemyslaw Prusinkiewicz [University of Calgary, Canada])

  Modeling natural elements such as trees in a plausible way, while offering simple and rapid user control, is a challenge. In a first collaboration with the EPI Imagine (ex-Evasion) we developed a method based on the design of plants from silhouettes [50]. This sketching paradigm allows quick and intuitive specification of foliage at multiple scales. On this topic, we started a collaboration with S. Longuay and P. Prusinkiewicz who develop iPad tools to design plants based on SCA. Combination of multitouch interface, sketching paradigm and powerfull adaptive procedural model that generate realistic trees offer intuitive and flexible design tools. This work is part of the Inria associated team with the University of Calgary. It has been published Eurographics Symposium on Sketch-Based Interfaces and Modeling [32].

- *Reconstruction from video.* (Frédéric Boudon, Jérome Guenard [IRIT, Toulouse], Géraldine Morin [IRIT, Toulouse], Pierre Gurdjos [IRIT, Toulouse], Vincent Charvillat [IRIT, Toulouse])

  Even if mature computer vision techniques allow the reconstruction of challenging 3D objects from images, due to high complexity of plant topology, dedicated methods for generating 3D plant models must be devised. We propose an analysis-by-synthesis method which generates 3D models of a plant from both images and a priori knowledge of the plant species.

  Our method is based on a skeletonisation algorithm that allows to generate a possible skeleton from a foliage segmentation. Then, a 3D generative model, based on a parametric model of branching systems that takes into account botanical knowledge is built. This method extends previous works by constraining the resulting skeleton to follow hierarchical organisation of natural branching structure.

3D models are then generated. A reprojection of the models can be compared with the original image to assess the visual accuracy of the reconstruction. We optimise the parameter values of the generative model based on the reprojection criterion. Realistic results are obtained on different species of plants, in particular vineyards. Publication of this work is in progress.

- *Reconstruction of virtual fruits from pictures.* (Mik Cieslak, Nadia Bertin [Inra, Avignon], Frédéric Boudon, Christophe Godin, Michel Genard [Inra, Avignon], Christophe Goz-Bac [Université Montpellier 2])

  This research theme is supported by the Agropolis project Fruit3D.

  The aim of this work is to provide methods for generating fruit structure that can be integrated with models of fruit function and used to investigate such effects. To this end, we have developed a modeling pipeline in the OpenAlea platform that involves two steps: (1) generating a 3D volumetric mesh representation of the entire fruit, and (2) generating a complex network of vasculature that is embedded within this mesh. To create the 3D volumetric mesh, we use reconstruction algorithms from the 3D mesh generation package of the Computational Geometry Algorithms Library (CGAL). To generate the pattern of vasculature within this volumetric mesh, we use a Space Colonisation Algorithm that populates the volume of the fruit by simulating competition for space of the vasculature. We have applied our modeling pipeline to generate the internal and external geometry of a cherry tomato fruit using Magnetic Resonance Imaging data as input. These studies demonstrate the possibility to create species-specific models of fruit structure with relatively low effort [26]. These volumetric meshes are then combined with models of function to form integrative computational fruit models, which will help to investigate the effects of fruit structure on quality (see section 5.3.2).

- *Reconstruction of gramineous leaves.* (Christian Fournier, Christophe Pradal)

  This research theme is supported by the Agropolis project OpenAlea.

  Unlike trees, the 3D architecture of gramineous plants is much more related to the shapes of its leaves than the arrangement of its branches. Many modeling efforts have thus concentrated on correctly capturing its complex shape at different stages and use them as scalable geometric primitives. Still, additional control of such objects is needed in the context of Functional Structural Modeling. The objective of this work is to propose a plastic and dynamic 3D leaf model that is well suited for such uses, still able to capture a variety of observed static shapes. Leaf shape is modeled by a parametric surface describing leaf midrib curvature, leaf width variation, undulation of leaf margins and twist along the midrib. Meshes can be generated form these surfaces, and reduced using a decimation algorithm. The model can be fed with data or with curves drawn by user interaction. Morphological operators are defined and allows for plastic deformation of the control curves. The dynamics of shape acquisition can also be specified, and combined with morphological operators to simulate various scenarii of evolution and responses to stresses. The capabilities of the model are demonstrated through several cases of use. Future directions of research are thought to be a better integration of mechanical or physiological constraints that would reduce the model plasticity but avoid user-induced unrealistic simulation. [28].

5.1.2. Modeling the plant ontogenic program

Participants: Christophe Godin, Yann Guédon, Evelyne Costes, Jean-Baptiste Durand, Anaëlle ambreville, Pierre Fernique, Christophe Pradal, Jean Peyhardi, Catherine Trottier, Yassin Refahi, Etienne Farcot.

This research theme is supported by two PhD programs.

The remarkable organization of plants at macroscopic scales may be used to infer particular aspects of meristem functioning. The fact that plants are made up of the repetition of many similar components at different scales, and the presence of morphological gradients, e.g. [37], [43], [44], [41], provides macroscopic evidence for the existence of regularities and identities in processes that drive meristem activity at microscopic scales. Different concepts have been proposed to explain these specific organizations such as "morphogenetic program" [45], "age state" [40] or "physiological age" [38]. All these concepts state that meristem fate changes according to position within the plant structure and during its development. Even though these changes in meristem fate are specific to each species and lead to the differentiation of axes, general rules can be highlighted [40], [38]. Here we develop computational methods to decipher these rules.

- *Relating branching structure to the shoot properties* (Jean Peyhardi, Yann Guédon, Evelyne Coste, Catherine Trottier, Yves Caraglio [AMAP], Pierre-Eric Lauri [AGAP, AFEF team]) Shoot branching structures often take the form of a succession of homogeneous branching zones and have been analyzed using segmentation models such as hidden semi-Markov chains. Axillary meristem fates are influenced by local properties of the parent shoot such as for instance its growth rate or local curvature. The objective of this work, which is part of the PhD subject of Jean Peyhardi, is to develop statistical models that generalize hidden semi-Markov chains with the capability to incorporate explanatory variables that vary along the parent shoot (e.g. maximum growth rate of the leaf, surface of the leaf, length of the internode, local curvature of the parent shoot). More precisely, the simple multinomial distributions that represent the axillary productions observed in the different branching zones are replaced by multinomial generalized linear models (GLMs). Since the two classical categories of multinomial GLMs that correspond either to nominal or ordinal categorical response variables were not appropriate in our setting, we chose to develop a new family of multinomial GLMs called multi-step multinomial GLMs that enable to tackle partially ordered categorical response variables. Typically, we need to distinguish different timing of branching (e.g. immediate shoot, one-year-delayed shoot and latent bud), different categories of offspring shoots (e.g. among one-year-delayed shoots, vegetative short shoot, vegetative long shoot and flowering shoot) and to specialize the explanatory variables for certain categories of offspring shoots (e.g. the growth of the parent shoot influence the immediate offspring shoots but not the one-year-delayed offspring shoots). The resulting integrative models are called semi-Markov switching multi-step multinomial GLMs and are applied to different data sets corresponding mainly to fruit tree branching structures.
- *Genetic determinisms of the alternation of flowering in apple tree progenies.* (Jean-Baptiste Durand, Jean Peyhardi, Baptiste Guitton [AGAP, AFEF team], Yan Holtz [AGAP, AFEF team] Catherine Trottier, Evelyne Costes, Yann Guédon)

  The aim of this work was to characterize genetic determinisms of the alternation of flowering in apple tree progenies. Data were collected at two scales: at whole tree scale (with annual time step) and a local scale (annual shoot or AS, which is the portions of stem that were grown during the same year). Two replications of each genotype were available.

  Indices were proposed for early detection of alternation during the juvenile phase. They were based on a trend model and a quantification of the deviation amplitudes and dependency, with respect to the trend. This allows for quantifying alternation from the yearly numbers of inflorescences at tree scale.

  However, phenotyping subsamples of AS sequences is more realistic in the framework of genotype selection. To model alternation of flowering at AS scale, a second-order Markov tree model was built. Its transition probabilities were modeled as generalized linear mixed models, to incorporate the effects of genotypes, year and memory of flowering for the Markovian part, with interactions between these components. Asynchronism of flowering at AS scale was also assessed using an entropy-based criterion.

  This work started during the PhD's work of Baptiste Guitton. It was then extended in 2012 by Yan Holtz during this Master 2 internship, supervised by Evelyne Costes and Jean-Baptiste Durand. New progenies were considered, as well as the performance of approximating the descriptors at whole tree scale with those at AS scale. These descriptors allowed the identification of QTL zones involved in the control of flowering in apple trees.

  As a perspective of this work, patterns in the production of children ASs (numbers of flowering and vegetative children) depending on the type of the parent AS must be analyzed using branching processes and different types of Markov trees, in the context of Pierre Fernique's PhD Thesis (see next item in Section 5.1.2).

- *Modeling branching patterns in fruit tree shoots through the characterization of their demographic properties* (Pierre Fernique, Jean-Baptiste Durand, Yann Guédon).

  To test the effect of some properties of a given parent shoot on the properties of its children shoots, statistical models based on multitype branching processes were developed. This kind of dependence between parent and children shoots is frequently at stake in fruit trees, for which the number of flowering or vegetative children of a parent shoot depends on its nature, with potential interactions with other factors. Thus, controlling demographic patterns of the shoots (through varietal selection or crop management strategies) is expected to bring substantial improvements in the quantity and quality of yields.

  Formally, the shoot properties are summed up using the notion of shoot state. The number of children shoots in each state is modeled through discrete multivariate distributions. Model selection procedures are necessary to specify parsimonious distributions. We developed an approach based on probabilistic graphical models to identify and exploit properties of conditional independence between numbers of children in different states, so as to simplify the specification of their joint distribution. The graph building stage was based on exploring the space of possible chain graph models, which required defining a notion of neighbourhood of these graphs. A parametric distribution was associated with each graph. It was obtained by combining families of univariate and multivariate distributions or regression models. These were chosen by selection model procedures among different parametric families.

  This work was carried out in the context of Pierre Fernique's first year of PhD (Montpellier 2 University and CIRAD). It was applied to model dependencies between short or long, vegetative or flowering shoots in apple trees. The results highlighted contrasted patterns related to the parent shoot state, with interpretation in terms of alternation of flowering (see previous item in Section 5.1.2). It was also applied to the analysis of the connections between cyclic growth and flowering of mango trees. This work will be continued during Pierre Fernique's PhD thesis, with extensions to other fruit tree species and other parametric discrete multivariate families of distributions, including covariates and mixed effects.

- *Analyzing fruit tree phenology* (Anaëlle Dambreville, Jean-Baptiste Durand, Pierre Fernique, Yann Guédon, Christophe Pradal, Pierre-Eric Lauri [AFEF team, AGAP], Frédéric Normand, Catherine Trottier) Mango is a tropical tree characterized by strong asynchronisms within and between trees. Causation networks explaining the vegetative and reproductive growths within and between growing cycles were studied on the basis of generalized linear models. We highlighted in this way marked interplays between structural and temporal components of tree structure development at three scales. At growth unit scale, a growth unit appeared early in the growing cycle had higher rate of burst compared to late appeared growth units. At growing cycle scale, a growth unit which flowered delayed its future vegetative growth compared to a vegetative growth unit. At tree scale, a fruiting tree delayed further vegetative growth and flowering compared to a non-fruiting tree. These results evidenced that tree phenology is strongly affected by structural components and not only by the environment. We are now investigating jointly structure development and phenology of mango using statistical models for trees in particular hidden Markov tree models and multitype branching processes.

- *Integrative developmental growth stages of shoots* (Anaëlle Dambreville, Yann Guédon, Pierre-Eric Lauri [AFEF team, AGAP], Frédéric Normand) Growth and development are often studied as two separated processes. Our aim is to investigate the coordination between growth and development in mango shoots. We considered three types of organ, namely the shoot axis, its attached leaves and the inflorescence. Two types of data were collected during the shoot and inflorescence follow-up: developmental stages determined in an expert way and organs sizes determined from measurements. To give an integrative view of the shoot and inflorescence growth and development, we adopted the following strategy. For a given cultivar, we first built a multi-state model on the basis of absolute growth rate sequences deduced from the measurements. Using these models, we computed growth stages. These growth stages highlighted growth asynchronisms between two topologically-connected organs: the axis and its leaves. Then, we compared these growth stages with the developmental ones and we obtained strong matches between them. The integrated developmental growth stages emphasized that the developmental stages are markedly related to growth rates an can be interpreted in terms of physiological (hydraulics, carbohydrates partitioning) and developmental (organs preformation versus neoformation) processes.

- *Self-nested structure of plants.* (Christophe Godin, Farah Ben Naoum) In a previous work [6], we designed a method to compress tree structures and to quantify their degree of self-nestedness. This method is based on the detection of isomorphic subtrees in a given tree and on the construction of a DAG, equivalent to the original tree, where a given subtree class is represented only once (compression is based on the suppression of structural redundancies in the original tree). In the compressed graph, every node representing a particular subtree in the original tree has exactly the same height as its corresponding node in the original tree. This method thus compresses a tree in width, but not in height. In this new work, we designed an extension of this compression method in which a tree is compressed in both width and height. The method is based on the detection of so-called *quasi-isomorphic paths* in a tree and on the compression of these paths in height. A paper describing the corresponding algorithms is being written.

5.1.3. Analyzing the influence of the environment on the plant ontogenic program

Participants: Frédéric Boudon, Jean-Baptiste Durand, Christophe Godin, Yann Guédon, Jean Peyhardi, Pierre Fernique, Maryline Lièvre, Christine Granier, Evelyne Costes, Pascal Ferraro, Catherine Trottier.

This research theme is supported by three PhD programs.

The ontogenetic program of a plant is actually sensitive to environmental changes. If, in particular cases, we can make the assumption that the environment is a fixed control variable (see section 5.1.2), in general the structure produced by meristem results from a tight interaction between the plant and its environment, throughout its lifetime. Based on observations, we thus aim to trace back to the different components of the growth (ontogenetic development and its modulation by the environment). This is made using two types of approaches. On the one hand, we develop a statistical approach in which stochastic models are augmented with additional time-varying explanatory variables that represent the environment variations. The design of estimation procedures for these models make it possible to separate the plant ontogenetic program from its modulation by the environment. On the other hand, we build reactive models that make it possible to simulate in a mechanistic way the interaction between the plant development and its environment.

- *Influence of environment conditions and horticultural practices on the branching and axillary flowering structures of fruit tree shoots.* (Yann Guédon, Evelyne Costes [AFEF Team, AGAP], Ted DeJong [UC Davis], Claudia Negron [UC Davis]).

  In the context of a collaboration with Claudia Negron and Ted DeJong, we studied the influence of water availability and pruning practices on the branching and axillary flowering structures of different categories of almond shoots Stochastic models (hidden semi-Markov chains) were built for the branching and axillary flowering structures of different categories of almond shoots corresponding to different genetic backgrounds, levels of irrigation and pruning practices.

- *Analyzing growth components in trees.* (Yann Guédon, Yves Caraglio [AMAP], Olivier Taugourdeau [AMAP])

  In a forest ecology context, we identified robust indicators that summarize the balance between tree ontogeny and environmental constraints (mainly related to light environment). In this context, tree growth data typically correspond to the retrospective measurement of annual shoot characteristics (e.g. length, number of branches) along the main stem. We applied segmentation models (hidden Markov and semi-Markov chains) that enable to identify tree growth phases. This statistical modeling approach was applied to both deciduous (sessile oak and Persian walnut) and evergreen (Corsican pine and silver fir) tree species growing in contrasted conditions ranging from managed forest stands to unmanaged understoreys. The growth phase duration distributions estimated within these segmentation models characterize the balance between tree ontogeny and the environmental constraints in tree development at the population scale. These distributions had very contrasted characteristics in terms of shape and relative dispersion between ontogeny-driven and environment-driven tree development. The characteristics of growth phase duration distributions may change over tree life reflecting changes in tree competition.

- *Investigating the influence of geometrical traits on light interception efficiency of trees and grass* (Liqi Han [AFEF Team, AGAP], Christophe Pradal, Frédéric Boudon, Christophe Godin, David Da Silva [UC Davis], Evelyne Costes, Philippe Balandier [PIAF], André Marquier [PIAF], Gaëtan Louarn [URP3F], Didier Combes [URP3F], Christian Fournier)

  Light availability in forest understory is essential for many processes. It controls for instance the growth potential of species and individuals in plant communities. It is, therefore, a valuable information regarding forest and crop management. However, the effects of competition for light on short term vegetation dynamics are still poorly understood. This is in part due to a lack of tractacle and precise methods to estimate light resource within a canopy. To alleviate this difficulty, models can be used to compute light interception. At a detailed scale, they often require a lot of field data to accurately predict light distribution, particularly in the case of heterogeneous canopies.

  To investigate this issue, we first analyzed the deterioration of the prediction quality of light distribution to the reduction of inputs by comparing simulations to transmitted light measurements in forests of increasing complexity in three different locations [17]. With a full set of parameters to describe the tree crown (i.e., crown extension in at least eight directions, crown height and length), the model accurately simulated the light distribution. Simplifying crown description by a geometric shape with a mean radius of crown extension led to deteriorated but acceptable light distributions. Allometric relationships used to calculate crown extension from trunk diameter at breast height seriously reduced light distribution accuracy.

  We also studied the light interception of herbaceous plants with contrasting architectures (monocultures and binary mixtures) grown at high or low density and sought to determine the important architectural features necessary to account for light partitioning among individual plants [21]. It was shown that the studied plant populations were typical of a wide range of competition intensities, ranging from sparse plants to dense size-structured populations. Plant representations using whole plant envelopes with homogeneous leaf area density (LAD) were not reliable to estimate light partitioning, irrespective of the accuracy of envelope definition. Accounting for heterogeneous LAD within plants helped to solve this problem in both sparse and dense canopies. The relative importance of traits however changed with competition intensity and was different from reports made on isolated plants. Simple envelope-based reconstructions were finally shown robust enough to support parameterisation from a tractable set of traits measured in the field provided that height and vertical LAD gradient were characterised.

  Using virtual growth simulation tools, a detailled analysis was also carried out to study more precisely the influence of architectural variability of apple trees on their light interception efficiency [30]. For this we used MAppleT, an in silico functional-structural plant model that has been built for simulating architectural development of apple trees. The STAR, namely the silhouette to total area ratio, of leaves, was chosen to evaluate the level of such efficiency. The strategy was to integrate MAppleT with the light interception model provided by the fractalysis module of the VPlants software library. Target values of four major traits (internode length, leaf area, branching angle and top shoot diameter), are varied in range previously observed in a segregating population of apple hybrids. A sensitivity analysis based on polynomial and generalised additive models was performed for highlighting the most influential trait on light interception and suggesting the optimal combination(s) of traits leading to the highest STAR. The contribution of stochastic processes that pilot tree topology in MAppleT is also investigated in the sensitivity analysis. This study not only provides a time- and resource-saving alternative for data collection, but also sets a methodology for ideotype definition and further genetic improvement of apple trees.

5.2. Meristem functioning and development

In axis 2 work focuses on the creation of a *virtual meristem*, at cell resolution, able to integrate the recent results in developmental biology and to simulate the feedback loops between physiology and growth. The approach is subdivided into several sub-areas of research.

5.2.1. Data acquisition and design of meristem models

Participants: Frédéric Boudon, Christophe Godin, Christophe Pradal, Vincent Mirabet [RDP, ENS], Jan Traas, Grégoire Malandain, Jean-Luc Verdeil [PHIV, AGAP].

This research theme is supported by the iSam and Morphogenetics projects.

- *Improvement of the MARS-ALT pipeline robustness* Meristem, laser microscopy, image reconstruction, cell segmentation, automatic lineaging
  Participants: Léo Guignard, Christophe Godin, Grégoire Malandain, Jan Traas, Pradeep Das [RDP, ENS], Vincent Mirabet [RDP, ENS].

  The MARS-ALT (Multi-Angles Registration and Segmentation - Automatic Lineage Tracking) software pipeline automatically performs a segmentation at cell resolution from 3D or 2D voxel images where the membranes/walls are marked (by a die for example) and makes it possible to follow the lineage of these cells through time [5]. A new version of this pipeline is currently being developed. MARS-ALT Version 2 is based on the same algorithms and methods and is intended to improve the overall robustness of the pipeline (protocol, noise in the input image) and automate completely the process. To test the new pipeline, we use different acquisition protocols and different organisms (floral and apical meristems and the early stages of development of a marine animal *Phallusia mammillata*). The segmentation is corrected a posteriori to deal with imaging artifacts due to uncertainties of acquisition. The image data set on which we develop the methods consists of :

  - *Arabidopsis thaliana* shoot apical meristem and primordia with around 6000 cells. The organ is captured from three different angles every 20 hours 3 or 4 times with a confocal microscope (Collaboration Sainsbury lab, Cambridge)
  - *Arabidopsis thaliana* flower meristems with around 2000 cells. The organ is also captured from three different angles every 20 hours 3 to 5 times with a confocal microscope (Collaboration RDP Lyon)
  - *Phallusia mammillata* and *Ciona intestinalis* embryos with from 32 cells to around 1000 cells. The organism is captured from four different angles every 2 minutes during 2 to 3 hours with a SPIM (Single Plane Illumination Microscope) (Collaboration CRBM Montpellier / EMBL Heidelberg)

  The pipeline provides as an output segmented images on which metrics for each cells can be extracted such as volume, principal components, convex hull and so on. A new non-linear registration algorithm developed by G. Malandain (MORPHEME team, Inria Sophia-Antipolis) is now available and will lead to an improvement of ALT algorithm. Redesign and improvement of the lineage tracking pipeline will be the next step.

- *Design of 3D virtual atlases for specifying gene expression patterns* (Jérôme Chopard, Christophe Godin, Jan Traas, Françoise Monéger [RDP, ENS])

  This research theme is supported the ANR GeneShape and iSam projects.

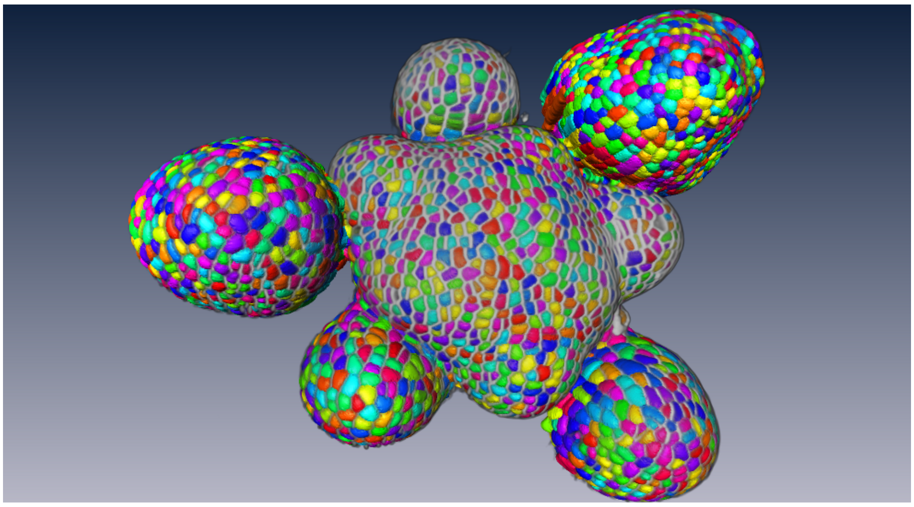


Figure 4. Superimposition of an automatic cell segmentation of an arabidopsis flower meristem using the new MARS pipeline with the original confocal image stack where the membranes are marked.

To organize the various genetic, physiological, physical, temporal and positional informations, we build a spatialized and dynamic database. This database makes it possible to store all the collected information on a virtual 3D structure representing a typical organ. Each piece of information has to be located spatially and temporally in the database. Tools to visually retrieve and manipulate the information, quantitatively through space and time are being developed. For this, the 3D structure of a typical organ has been created at the different stages of development of the flower bud. This virtual structure contains spatial and temporal information on mean cell numbers, cell size, cell lineages, possible cell polarization (transporters, microtubules), and gene expression patterns. Such 3D virtual atlas is mainly descriptive. However, like for classical databases, specific tools make it possible to explore the virtual atlas according to main index keys, in particular spatial and temporal keys. Both a dedicated language and a 3D user interface are being designed to investigate and query the 3D virtual atlas.

A prototype version of the 3D virtual atlas was built last year [8]. Further developments of this tool will rely on the segmented images produced from microscopy, as presented in the previous section. In particular, a common underlying data structure has to be developed transversally to these two scientific developments. The definition of this data structure has been initiated last year through several team meetings, and should lead to a revised implementation next year.

5.2.2. Shape analysis of meristems

(Jonathan Legrand, Clémence Hatt [BURST, AGAP], Jean-Baptiste Durand, Frédéric Boudon, Christophe Godin, Yann Guédon, François Mankessi [BURST, AGAP], Olivier Monteuuis [BURST, AGAP], Jean-Luc Verdeil [PHIV, AGAP])

Plants that grow several forms or type of leaves along a shoot, depending on age or shoot length, are called heteroblastic.The influence of heteroblasty on morphological and histocytological characteristics of Acacia mangium shoot apical meristems (SAMs) was assessed comparing materials with mature and juvenile leaf morphology in natural and in vitro conditions. For this we introduced a workflow for characterizing dome shape with few parameters (SAM dome heigth (H), basal diameter (D) and shape factor (S)) and their joint statistical analysis to assess influence of conditions on SAM shape. In particular, a new statistical test is introduced here for multivariate analysis. This is a generalization of univariate ANOVA that takes into account statistical dependencies between the shape parameters. As a result, we found that SAM dome height (H) and basal diameter (D) were highly correlated. The joint analysis revealed that H, D, and shape (S) varied significantly according to the four plant origins investigated, with the higher scores for the outdoor mature source "Mat". Overall, heteroblasty induced more conspicuous differences of SAM characteristics for the outdoor than for the in vitro materials. A paper presenting these results has been published in Trees [20]

At cellular resolution, we studied the organization of cells in the meristems. The MARS-ALT pipeline provides rich spatio-temporal data sets for analyzing the development of meristems. A first step consisted of designing a dedicated graph for efficiently representing the spatial (adjacency between cells) and temporal (cell division) relationships between cells. Various variables can be attached either to the vertices (e.g. cell volume, inertia axes) or the edges (e.g. wall surface, distance between cell centroids). This graph may be augmented by new variables resulting from various spatial or temporal filtering (e.g. cell volumetric growth). We are now designing models and algorithms for finding patterns in time courses of meristems. In particular, we are investigating spectral clustering methods in order to define homogeneous regions in term of cell identities.

5.2.3. Transport models

Participant: Michael Walker.

This research theme is supported by the ANR GeneShape and ERASysBio+ iSAM projects and Morphogenetics.

Active transport of the plant hormone auxin has been shown to play a key role in the initiation of organs at the shoot apex, and vein formation in both leaves and the shoot apical meristem. Polar localized membrane proteins of the PIN1 and AUX/LAX family facilitate this transport and observations and models suggest that the coherent organization of these proteins in the L1 layer is responsible for the creation of auxin maxima (surrounded by a depletion zone), which in turn triggers organ initiation close to the meristem center [46] [1]. Furthermore, canalized PIN allocations are thought to play a crucial role in vein formation in the leaf and in the L2. Previous studies have typically modeled the L1 and L2 with different models to explain different patterns of PIN allocations. In the last two years, we developed a unifying model showing that a unique flux-based model could be sufficient to explain PIN patterns in both L1 and L2 [25]. Contrary to our previous study [11], here no change in the model parameters is needed for this. Our approach is based on inherent topological and geometrical differences between the L1 and L2, specifically their dimensionality and the distribution of sources and sinks.

In a different perspective, another study on auxin transport models have been submitted this year. In this work, a generic, adimensional flux-based model of auxin transport was studied using a combination of analytic and numeric approach. The steady-states with uniform auxin distribution where characterised for arbitrary tissues, and some of their bifurcations (loss of stability and Hopf) were described. This work, initiated during an "Explorateur" project funded by Inria during the period October 2012-January 2013, was submitted by E. Farcot and Y. Yuan (Memorial University of Newfoundland, Canada) in September and is still under review.

5.2.4. Mechanical model

Participants: Jérôme Chopard, Olivier Ali, Christophe Godin, Frédéric Boudon, Jan Traas, Olivier Hamant [ENS-Lyon], Arezki Boudaoud [ENS-Lyon].

This research theme is supported by the ANR VirtualFlower and Geneshape projects together with the Inria project Morphogenetics and the ERC from Jan Traas.

The rigid cell walls that surround plant cells is responsible for their shape. These structures are under constraint due to turgor pressure inside the cell. To study the changes of shape in plant tissues during organogenesis, we need a mechanical model of tissue development at cellular resolution. We developed such a model, in which walls are characterized by their mechanical properties like the Young modulus which describes the elasticity of the material. Wall deformation results from forces due to turgor pressure. Growth results from cell wall synthesis that is triggered when wall deformation exceeds a particular threshold. The final shape of the tissue integrates mechanically all the local deformations of each cell.

To model this process, we used a tensorial approach to describe both tissue deformation and stresses. Deformations were decomposed into elementary transformations that can be related to underlying biological processes. However, we showed that the observed deformations does not map directly local growth instructions given by genes and physiology in each cell. Instead, the growth is a two-stage process where genes are specifying how cell walls should yield to mechanical stresses. In this way, different regions in the tissue with different cell identities can have different growth properties. The final shape of the tissue results from the integration of all these mechanical properties and stresses at organ level under the growth force due to turgor pressure at tissue scale.

A paper describing the mechanical model and its application to model primorium formation in the shoot apical meristem is currently being written. Additionally, a redesign of our mechanical model using the SOFA framework is in progress.

5.2.5. Gene regulatory networks

Modeling gene activities within cells is of primary importance since cell identities correspond to stable combination of gene expression. This fact is becoming more and more acknowledged, and has led this year to the publication of several review articles including members of Virtual Plants as authors [22], [23], [24], [13].

- *The auxin signaling pathway* (Etienne Farcot, Yann Guédon, Christophe Godin, Yassin Refahi, Jonathan Legrand, Jan Traas, Teva Vernoux, Stéphane Robin [AgroParisTech], Jean-Benoist Leger [AgroParisTech])

  The auxin signalling network involves about 50 potentially interacting factors. We applied a graph clustering method [12] that relies on 0/1 interactions between factors deduced from yeast two-hybrid (Y2H) data. The Y2H analysis involves two independent tests (X-gal and HIS3 tests). Each possible interaction was tested in the two possible configurations, where each protein was alternatively the bait and the prey protein. A binary interaction is thus a summary of the four outputs of the X-gal and HIS3 tests. In order to limit the loss of information, we designed a standardization procedure to summarize the outputs of the X-gal and HIS3 tests as a distance defined on a continuous scale. This opens the possibility to study the influence of phylogenetic distances between factors on their interactions using an extension of the mixture model for random graphs that incorporate explanatory variables. This new model evidences different behaviors between the ARF+ and the Aux/IAA factors.

  Extensions of this ODE model are necessary to better understand this system in more general contexts than the shoot apical meristem development, which was the framework of our previous study. This work involves defining and studying mathematically a series of distinct network topologies for the auxin signalling pathway. This was the topic of a Master's internship for the studend Cyril Lavedrine, from September to December 2012, co-supervised by E. Farcot and T. Vernoux. This work led to qualitative predictions which remain to be tested experimentally.

- *Complex dynamics and spatial interactions in gene networks* (Yassin Refahi, Etienne Farcot, Christophe Godin)

  Complex computational and mathematical questions arise in the study of gene networks at two levels: (i) the single cell level, due to complex, nonlinear interactions, (ii) the tissue level, where multiple cells interact through molecular signals and growth, so that even simple local rules can challenge our intuition at higher scales.
  At the single cell level, new results were obtained in the framework of piecewise-linear models. Since their introduction in the late 1960's, these models have been believed to present chaotic behavior in some parameter regimes. However, this was mostly observed numerically, based on intensive generation of random networks. In a long lasting collaboration between E. Farcot and R. Edwards (Univ. Victoria, Canada), with more recent input from one of his students, E. Foxall, we have introduced a method to explicitly build piecewise affine models having a return map which is conjugate to a topological horseshoe. A paper presenting these results has appeared this year [18].

For the same class of piecewise-linear models, it is in general very difficult to entirely characterize the attractors of a given system. In an attempt to improve our ability on this question, a probabilistic approach has been proposed in [15], in which it is shown that a Markov chain can built as an approximation of a given piecewise-linear system, and actually used to make predictions about its periodic attractors.

At a higher scale, we have also continued the study of gene regulation in meristematic tissues. In the context of Y. Refahi's post-doc between Virtual Plants and the group of Henrik Jönsson in Cambridge (Sainsbury Laboratory), we have continued a work that was initiated in Y. Refahi's thesis. This work is motivated by recent biological results, indicating that gradient-like patterns originating from the external layers of meristems may play a decisive role in the specification of the pool of stem cells in a central position. Using the methods in [5], and their on-going improvements, we have acquired new 3D and 4D images that were then segmented. These structures will be used in the next few months to investigate generic patterning properties of gradient like morphogen patterns. This will require a thorough analysis of free diffusion in realistic geometries, as made possible by the newly acquired images. As a preliminary work, we are also currently investigating the formation of gradient patterns in idealised tissues, allowing for deeper analytic treatment than the complex structures obtained by microscopy.

5.2.6. Model integration

Participants: Mikaël Lucas [IRD], Michael Walker, Jérôme Chopard, Frédéric Boudon, Christophe Godin, Laurent Laplaze, Jan Traas, François Parcy.

This research theme is supported by the ANR/BBSRC project iSam.

Our approach consists of building a programmable tissue which is able to accept different modeling components. This includes a central data structure representing the tissue in either 2-D or 3-D, which is able to grow in time, models of gene activity and regulation, models of signal exchange (physical and chemical) between cells and models of cell cycle (which includes cell division). For each modeling component, one or several approaches are investigated in depth, possibly at different temporal and spatial scales, using the data available from the partners (imaging, gene networks, and expression patterns). Approaches are compared and assessed on the same data. The objective of each submodel component will be to provide plugin components, corresponding to simplified versions of their models if necessary, that can be injected in the programmable tissue platform. This work is developed in collaboration with the RDP grou at ENS-Lyon [23] and the CPIB group in Nottingham, UK [13].

- *Development of a computer platform for the 'programmable tissue'.* (Michael Walker, Frédéric Boudon, Etienne Farcot, Christophe Godin)

  One key aspect of our approach is the development of a computer platform dedicated to programming virtual tissue development. This platform will be used to carry out integration of the different models developed in this research axis. The platform is based on *OpenAlea*. Partner models can be integrated in the platform in a non-intrusive way (the code of their model need not be rewritten). In this context, model integration will i) consist of designing adequate data-structures at different levels that will be exchanged and reused among the different plug-in models and ii) defining control flows at adequate levels to avoid the burden of excessive interaction between components. In the past year, progress has been made in defining a generic tissue data structure that could be used in this platform, through several group meetings along the year. A redesign of the structure is in progress.

- *Design of a genetic model of inflorescence development.* (Etienne Farcot, Christophe Godin, François Parcy)

  We studied the regulatory network that control flower development during morphogenesis. To overcome the network complexity and integrate this regulation during ontogenesis, we have developed a first model of the control of floral initiation by genes, and in particular the situation of cauliflower mutants, in which the repeatedly meristem fails in making a complete transition to the flower. This work couples models at different scales, since gene regulation is described by a minimal gene network, which is used as a decision module in an L-system model of the inflorescence architecture. This mixed model has led us to make different hypotheses about gene interactions and hormonal regulation. First predictions about gene actors controling the passage to flower could be verified. However, a complete integrated picture of flower development could not be reached yet. After several unsuccessful attempts, further experiments are currently being made to verify the scenario predicted by the model.

5.3. Multi-scale models and analysis: from cells to plant architecture (and back)

5.3.1. Transport model in roots

Participants: Mikaël Lucas [IRD], Christophe Pradal, Christophe Godin, Christophe Maurel [BPMP].

This research theme is supported by the ANR project HydroRoot.

A model of Arabidopsis thaliana root hydraulics at the cellular level was developped in the OpenAlea modeling platform. The model relies on the integration throughout root architecture of elementary hydraulic components. Each component integrates local radial and axial water flows. Axial hydraulic conductivity is calculated according to Poiseuille's law, based on local size of xylem vessels. Radial hydraulic conductivity is determined in part by aquaporin activity and was set constant throughout root architecture in the first model versions. In its current state, the model is parameterized using architectural, tissular and physiological data that were experimentally determined in the Aquaporin group at BPMP. The architectural reconstruction of the root system is based on a tridimensional multi-scale tree graph (MTG). The current model is capable of predicting the water flow that is transported by a root system in the standard experimental conditions used in the Aquaporin group. This model was used to perform sensitivity analyses and determine the respective contributions to root hydraulic dynamics of various biological parameters (axial and radial hydraulic conductivites, root architecture). One major finding is that the root hydraulic conductivity (Lpr) computed from the model is highly dependent on root architecture. This is due to the limiting role of axial (xylem) conductance, one feature that had been neglected in previous representations of root water transport. The radial hydraulic conductivity may primarily be limiting in conditions of Lpr inhibition, since its increase from values in control roots has marginal effects on Lpr. A new set of experimental data including root diameter repartitions in wild-type plants, and xylem vessel diameters in mutants with altered xylem morphology (irx3, esk1) will be used to implement the model. Root cell hydraulic conductivities will also be measured in these and aquaporin mutant phenotypes. Our aim is to check whether, based on anatomical and morphological data, the model can properly predict the radial hydraulic conductivity of these genotypes.

5.3.2. Transport in fruits

Participants: Mik Cieslak, Nadia Bertin [Inra, Avignon], Frédéric Boudon, Christophe Godin, Michel Genard [Inra, Avignon], Christophe Goz-Bac [Université Montpellier 2].

This research theme is supported by the Agropolis project Fruit3D.

Understanding the controlling factors of fruit quality development is challenging, because fruit quality results from the interplay between physical and physiological processes that are under the control of genes and the environment. Although process-based models have been used to make significant progress in understanding these factors, they ignored to a large extent the shape and internal structure of the fruit.

Two essential functions in determining fruit quality are the transport and accumulation of water and dry matter to various fruit tissues. Since water and carbon are delivered to fruit tissues through a complex vasculature system, the internal fruit structure and pattern of vasculature may have a significant impact on their distribution within the fruit.

To help characterizing effects of fruit shape and internal structure on quality, the creation of a 3D virtual fruit model that integrates fruit structure and function with growth governed by environmental inputs has been investigated. For this, a modeling pipeline has been developed that includes the following steps: creation of a 3D volumetric mesh of the internal fruit structure, including vasculature (see section 4). Based on previous compartment models of fruit physiology developed at Avignon, we have then developed models of water and carbon transport that have been coupled with the 3D model of fruit. In the 3D model, different equations are describing the transport between adjacent regions of the fruit represented as a 3D mesh. The integration through space and time is carried out using a standard integration scheme (Runge-Kutta of order 4).

This approach has been applied to study tomato fruit (Solanum lycopersicum) by constructing 3D volumetric meshes from different sources (images of perpendicular fruit slices and MRI data), and integrating water and carbon transport processes into these meshes. To illustrate the tomato model, a simulation of one season of the fruit's growth has been performed and its results compared with an already published process-based tomato fruit model. We first showed that the classical results of the abstract process-based models could be also captured by the more detailed spatialized model. However, our model provides additional information on the internal heterogeneity of the fruit, such as a gradient in sugar concentration. Once the model is calibrated and evaluated, our approach will be suitable for studying the effects of internal fruit heterogeneity and overall shape on fruit quality development.

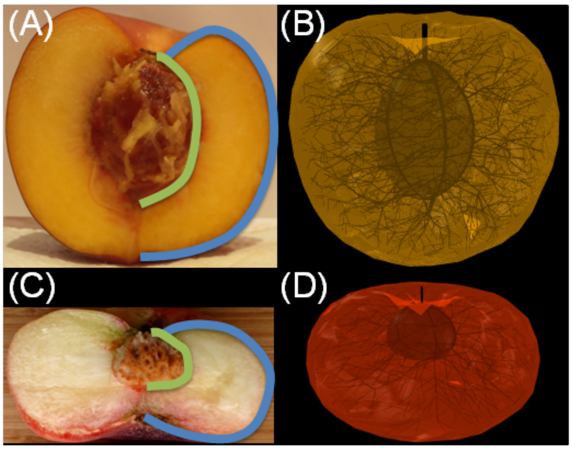


Figure 5. Virtual models of peaches reconstructed from images with simulated vasculatur to simulate carbon and water transport in the fruit

5.3.3. Analazing shoot and leaf elongation

Participants: Maryline Lièvre, Yann Guédon, Christine Granier.

The analysis of phenotyping data coming from automated platforms such as PHENOPSIS often focuses on the growth of a leaf at a given rank along the stem. We aim at developing a pipeline of methods for analyzing the growth of *arabidopsis* shoot at three scales:

1. tissular scale using a probabilistic model of endoreduplication for modeling the distribution of the leaf epidermis cell surfaces. Endoreduplication, which is a replication of the nuclear genome in the absence of cell division that leads to elevated nuclear gene content, strongly affects the leaf epidermis cells of *arabidopsis*.
2. organ scale using nonlinear regression model for analyzing the growth of each successive leaf.

3. shoot scale: The outputs of the analyses at the tissular and organ scales will be summarized as multivariate sequences along the shoots characterizing each successive leaf. These sequences will be augmented by supplementary morphological variables characterizing leaf shape and properties (e.g. presence/absence of trichomes). These sequences will be globally analyzed in order to take into account plant ontogeny and in particular the successive developmental stages before the floral transition for the wild type and selected mutants of *arabidopsis*.

5.3.4. Analyzing perturbations in Arabidopsis thaliana phyllotaxis

Participants: Christophe Godin, Yann Guédon, Yassin Refahi, Etienne Farcot.

This research theme is supported by iSAM.

The cytokinin hormones are known to play a significant role in the regulation of phyllotaxis. To investigate this, Fabrice Besnard and Teva Vernoux are studying *Arabidopsis thalianaahp6* mutants, AHP6 being a protein known for its inhibitory effect in the cytokinin signaling pathway. At the macroscopic scale, this mutation induces perturbations of the phyllotaxis, barely sensible on single plants. In order to characterize these perturbations, we designed a pipeline of models and methods which relies of combinatorial and statistical techniques. Using this pipeline of methods, we have shown that the perturbation patterns in both wild-type and mutant plants can be explained by permutations in the order of insertion along the stem of 2 or 3 consecutive organs. The number of successive synchronized organs between two permutations reveals unexpected patterns that depend on the nature of the preceding permutation (2- or 3-permutation). We identified significant individual deviations of the level of baseline segments with reference to 137.5°, which confirms theoretical model predictions. Finally, we highlighted a marked relationship between permutation of organs and defects in the elongation of the internodes in between these organs. All these results can be explained by the absence of a strict coupling between the timing of organ development and their angular and longitudinal position on the stem. Two papers (one with biological aspects and the other about methodological developments) are currently in revision.

VISAGES Project-Team

6. New Results

6.1. Image Segmentation, Registration and Analysis

6.1.1. Estimating A Reference Standard Segmentation with Spatially Varying Performance Parameters: Local MAP STAPLE

Participant: Olivier Commowick.

We present a new algorithm, called local MAP STAPLE, to estimate from a set of multi-label segmentations both a reference standard segmentation and spatially varying performance parameters. It is based on a sliding window technique to estimate the segmentation and the segmentation performance parameters for each input segmentation. In order to allow for optimal fusion from the small amount of data in each local region, and to account for the possibility of labels not being observed in a local region of some (or all) input segmentations, we introduce prior probabilities for the local performance parameters through a new maximum a posteriori formulation of STAPLE. Further, we propose an expression to compute confidence intervals in the estimated local performance parameters. We carried out several experiments with local MAP STAPLE to characterize its performance and value for local segmentation evaluation. First, with simulated segmentations with known reference standard segmentation and spatially varying performance, we show that local MAP STAPLE performs better than both STAPLE and majority voting. Then we present evaluations with data sets from clinical applications. These experiments demonstrate that spatial adaptivity in segmentation performance is an important property to capture. We compared the local MAP STAPLE segmentations to STAPLE, and to previously published fusion techniques and demonstrate the superiority of local MAP STAPLE over other state-of-the-art algorithms.

This work was done in collaboration with Alireza Akhondi-Asl and Simon K. Warfield [15].

6.1.2. Voxel-based quantitative analysis of brain images from F-18 Fluorodeoxyglucose Positron Emission Tomography with a Block-Matching algorithm for spatial normalization

Participant: Olivier Commowick.

Statistical Parametric Mapping (SPM) is widely used for the quantitative analysis of brain images from F-18 fluorodeoxyglucose positron emission tomography (FDG PET). SPM requires an initial step of spatial normalization to align all images to a standard anatomic model (the template), but this may lead to image distortion and artifacts, especially in cases of marked brain abnormalities. This study aimed at assessing a block-matching (BM) normalization algorithm, where most transformations are not directly computed on the overall brain volume but through small blocks, a principle that is likely to minimize artifacts. Large and/or small hypometabolic areas were artificially simulated in initially normal FDG PET images to compare the results provided by statistical tests computed after either SPM or BM normalization. Results were enhanced by BM, compared with SPM, with regard to (i) errors in the estimation of large defects volumes (about 2-fold lower) because of a lower image distortion, and (ii) rates of false-positive foci when numerous or extended abnormalities were simulated. These observations were strengthened by analyses of FDG PET examinations from epileptic patients. Results obtained with the BM normalization of brain FDG PET appear more precise and robust than with SPM normalization, especially in cases of numerous or extended abnormalities.

This work was done in collaboration with Christophe Person, Valérie Louis-Dorr, Sylvain Poussier, Grégoire Malandain, Louis Maillard, Didier Wolf, Nicolas Gilet, Véronique Roch, Gilles Karcher and Pierre-Yves Marie [19].

6.1.3. Block-matching strategies for rigid registration of multimodal medical images

Participants: Olivier Commowick, Sylvain Prima.

We propose and evaluate a new block-matching strategy for rigid-body registration of multimodal or multisequence medical images. The classical algorithm first matches points of both images by maximizing the iconic similarity of blocks of voxels around them, then estimates the rigid-body transformation best superposing these matched pairs of points, and iterates these two steps until convergence. In this formulation, only discrete translations are investigated in the block-matching step, which is likely to cause several problems, most notably a difficulty to tackle large rotations and to recover subvoxel transformations. We propose a solution to these two problems by replacing the original, computationally expensive, exhaustive search over translations by a more efficient optimization over rigid-body transformations. The optimal global transformation is then computed based on these local blockwise rigid-body transformations, and these two steps are iterated until convergence. We evaluate the accuracy, robustness, capture range and run time of this new block-matching algorithm on both synthetic and real MRI and PET data, demonstrating faster and better registration than the translation-based block-matching algorithm [28].

6.1.4. Automated diffeomorphic registration of anatomical structures with rigid parts: Application to dynamic cervical MRI

Participants: Olivier Commowick, Sylvain Prima.

We propose an iterative two-step method to compute a diffeomorphic non-rigid transformation between images of anatomical structures with rigid parts, without any user intervention or prior knowledge on the image intensities. First we compute spatially sparse, locally optimal rigid transformations between the two images using a new block matching strategy and an efficient numerical optimiser (BOBYQA). Then we derive a dense, regularised velocity field based on these local transformations using matrix logarithms and M-smoothing. These two steps are iterated until convergence and the final diffeomorphic transformation is defined as the exponential of the accumulated velocity field. We show our algorithm to outperform the state-of-the-art log-domain diffeomorphic demons method on dynamic cervical MRI data [27].

6.1.5. Computer-assisted paleoneurology

Participant: Sylvain Prima.

In collaboration with Antoine Balzeau and colleagues at the MNHN (http://www.mnhn.fr), we made the first ever description of the “digital” endocranial cast of the Cro-Magnon 1 specimen, discovered in 1868 at the Eyzies-de-Tayac, Dordogne, France [13]. Together with Benoit Combès (Géosciences Rennes, UMR 6118), we were especially involved in the assessment of its endocranial asymmetries, using an algorithm previously developed at VisAGeS [51] in the context of the ARC 3D-MORPHINE coordinated by Sylvain Prima (http://3dmorphine.inria.fr).

6.2. Image processing on Diffusion Weighted Magnetic Resonance Imaging

6.2.1. Non-Local Robust Detection of DTI White Matter Differences with Small Databases

Participants: Olivier Commowick, Aymeric Stamm.

Diffusion imaging, through the study of water diffusion, al- lows for the characterization of brain white matter, both at the population and individual level. In recent years, it has been employed to detect brain abnormalities in patients suffering from a disease, e.g. from multiple sclerosis (MS). State-of-the-art methods usually utilize a database of matched (age, sex, ...) controls, registered onto a template, to test for differences in the patient white matter. Such approaches however suffer from two main drawbacks. First, registration algorithms are prone to local errors, thereby degrading the comparison results. Second, the database needs to be large enough to obtain reliable results. However, in medical imaging, such large databases are hardly available. In this paper, we propose a new method that addresses these two issues. It relies on the search for samples in a local neighborhood of each pixel to increase the size of the database. Then, we propose a new test based on these samples to perform a voxelwise comparison of a patient image with respect to a population of controls. We demonstrate on simulated and real MS patient data how such a framework allows for an improved detection power and a better robustness and reproducibility, even with a small database [26].

6.2.2. Registration and Analysis of White Matter Group Differences with a Multi-Fiber Model

Participant: Olivier Commowick.

Diffusion magnetic resonance imaging has been used extensively to probe the white matter in vivo. Typically, the raw diffusion images are used to reconstruct a diffusion tensor image (DTI). The incapacity of DTI to represent crossing fibers leaded to the development of more sophisticated diffusion models. Among them, multi-fiber models represent each fiber bundle independently, allowing the direct extraction of diffusion features for population analysis. However, no method exists to properly register multi-fiber models, seriously limiting their use in group comparisons. This paper presents a registration and atlas construction method for multi-fiber models. The validity of the registration is demonstrated on a dataset of 45 subjects, including both healthy and unhealthy subjects. Morphometry analysis and tract-based statistics are then carried out, proving that multi-fiber models registration is better at detecting white matter local differences than single tensor registration.

This work was done in collaboration with Maxime Taquet, Benoit Scherrer, Jurriaan Peters, Mustafa Sahin, Benoît Macq and Simon K. Warfield [44].

6.2.3. Automated delineation of white matter fiber tracts with a multiple region-of-interest approach

Participant: Olivier Commowick.

White matter fiber bundles of the brain can be delineated by tractography utilizing multiple regions-of-interest (MROI) defined by anatomical landmarks. These MROI can be used to specify regions in which to seed, select, or reject tractography fibers. Manual identification of anatomical MROI enables the delineation of white matter fiber bundles, but requires considerable training to develop expertise, considerable time to carry out and suffers from unwanted inter- and intra-rater variability. In a study of 20 healthy volunteers, we compared three methodologies for automated delineation of the white matter fiber bundles. Using these methodologies, fiber bundle MROI for each volunteer were automatically generated. We assessed three strategies for inferring the automatic MROI utilizing nonrigid alignment of reference images and projection of template MROI. We assessed the bundle delineation error associated with alignment utilizing T1-weighted MRI, fractional anisotropy images, and full tensor images. We confirmed the smallest delineation error was achieved using the full tensor images. We then assessed three projection strategies for automatic determination of MROI in each volunteer. Quantitative comparisons were made using the root-mean-squared error observed between streamline density images constructed from fiber bundles identified automatically and by manually drawn MROI in the same subjects. We demonstrate that a multiple template consensus label fusion algorithm generated fiber bundles most consistent with the manual reference standard.

This work was done in collaboration with Ralph Suarez, Sanjay Prabhu and Simon K. Warfield [23].

6.2.4. Corticospinal tractography with morphological, functional and diffusion tensor MRI: a comparative study of four deterministic algorithms used in clinical routine

Participants: Sylvain Prima, Camille Maumet, Jean-Christophe Ferré.

In collaboration with Romuald Seizeur, Nicolas Wiest-Daesslé and Xavier Morandi, we aimed to compare four deterministic tractography algorithms used in clinical routine for the study of the corticospinal tract (the bundle mediating voluntary movement) in 15 right-handed volunteers. We found no difference between right and left sides of the brain for any of the algorithms [22].

6.2.5. A new multi-directional fiber model for low angular resolution diffusion imaging

Participants: Aymeric Stamm, Christian Barillot.

Diffusion MRI is a tool of choice for the analysis of the brain white matter fiber pathways. When translated to clinics, the short acquisition time leads to low angular resolution diffusion (LARD) images. Fiber pathways are then inferred assuming Gaussian diffusion (a.k.a. DTI) that provides one fiber orientation per voxel. In the past decade, recent researches highlight more intricate intra-voxel fiber configurations using higher angular resolution diffusion images. In collaboration with Patrick Perez (Technocolor), we have proposed a non-Gaussian diffusion model of the white matter fibers able to recover from crossing of fibers even from low angular resolution.This model enables crossing fibers to be theoretically estimated from only 8 diffusion MR images. In particular, this model allows for the retrospective study of DW data sets acquired in the past. [42] [43].

6.3. Medical Image Computing in Brain Pathologies

6.3.1. Detection of dysplasia and heterotopia

Participants: Elise Bannier, Camille Maumet, Jean-Christophe Ferré, Christian Barillot.

Focal cortical dysplasia and heterotopias are a recognized cause of epilepsy. Indication for surgery relies on precise localization and delineation. However, visual depiction of focal cortical dysplasia and heterotopias is difficult, time-consuming and reader dependant. Several 3D T1 voxel based morphometry methods have been proposed to automatically identify and suggest potential abnormalities to the reader. Several studies have shown the ability of Double Inversion Recovery imaging to detect intracortical lesions in MS and Epilepsy. In this study we propose to evaluate the ability of Double Inversion Recovery voxel based analysis to detect cortical and juxtacortical lesions in pharmaco resistant epileptic patients. This work was performed in collaboration with Arnaud Biraben, Anca Pasnicu and Eduardo Pasqualini, Béatrice Carsin-Nicol [24].

6.3.2. MRI Estimation of T_1 Relaxation Time Using a Constrained Optimization Algorithm

Participants: Fang Cao, Olivier Commowick, Elise Bannier, Jean-Christophe Ferré, Gilles Edan, Christian Barillot.

We propose a new method to improve T_1 mapping with respect to the popular *DESPOT1* algorithm. A distance function is defined to model the distance between the pure signal and the measurements in presence of noise. We use a constrained gradient descent optimization algorithm to iteratively find the optimal values of T_1 and M_0. The method is applied to MR images acquired with 2 gradient echo sequences and different flip angles. The performance of T_1 mapping is evaluated both on phantom and on in vivo experiments [25].

6.3.3. Characterization and Modeling of Multidimensional MRI signatures in Multiple Sclerosis in clinically isolated syndromes.

Participants: Yogesh Karpate, Olivier Commowick, Gilles Edan, Christian Barillot.

Clinically Isolated Syndrome data contribute to critical factors in obtaining meaningful precursor and predictors of Multiple Sclerosis. Current methodologies don’t go beyond segmentation and which generalize poorly over multi-modal MRI data. The project objective is to research and develop a framework for characterization and modeling of multidimensional MRI signatures in clinically isolated syndrome(disease’s onset),based on earlier and concurrent research and developments in the lab. In on going work an algorithmic framework is being developed to address the MS lesions’ classification ,identification and retrieval in USPIO-6 database.

As part of a battery of pre- processing techniques ,the module for intensity normalization of MRI volumes based on Spatio-Temporal Robust Expectation Maximization (STREM) is developed. This work is primarily based on 3 MRI modalities viz T1-w,T2-w and FLAIR. Complementary to this work , an another intensity normalization algorithm is devised based on parametric robust as well as efficient estimation by minimizing a density power divergence (beta divergence).The proposed method is indexed by a single parameter alpha which controls the trade-off between robustness and efficiency. The methodology affords a robust extension of maximum likelihood estimation for which alpha tends to be zero. Choices of alpha near zero afford considerable robustness while retaining efficiency close to that of maximum likelihood.

Moving forward, to facilitate accurate lesion tracking,features must be selected which are robust to photometric and geometric distortions. Energy measures are used to capture lesion's multiscale orientation structure in space.To illustrate utility with respect to a lesion detection, we have developed descriptor like local energy based on 3D steerable wavelets. This will be followed by the rigorous empirical evaluations of the resulting algorithm yielding better lesion identification and retrieval.

6.3.4. Multiple Sclerosis Lesions Evolution in Patients with Clinically Isolated Syndrome.

Participants: Alessandro Crimi, Olivier Commowick, Gilles Edan, Christian Barillot.

Multiple sclerosis (MS) is a disease with heterogeneous evolution among the patients. Some classifications have been carried out according to either the clinical course or the immunopathological profiles. Epidemiological data and imaging are showing that MS is a two-phase neurodegenerative in inflammatory disease. At the early stage it is dominated by focal in inflammation of the white matter (WM), and at a latter stage it is dominated by diffuse lesions of the grey matter and spinal cord. A Clinically Isolated Syndrome (CIS) is a first neurologic episode caused by in inflammation/demyelination in the central nervous system which may lead to MS. Few studies have been carried out so far about this initial stage. Better understanding of the disease at its onset will lead to a better discovery of pathogenic mechanisms, allowing suitable therapies at an early stage. We propose a new data processing framework able to provide an early characterization of CIS patients according to lesion patterns, and more specifically according to the nature of the inflammatory patterns of these lesions. Our method is based on a two layers unsupervised clustering. Initially, the spatio-temporal lesion patterns are classified using a tensor-like representation. The discovered lesion patterns are then used to identify group of patients and their correlation to one year follow-up total lesion loads, which is so far the only image-based figure that can potentially correlate to future evolution of the pathology. We expect that the proposed framework can infer new prospective figures from the earliest imaging sign of MS since it can provide a classification of different types of lesion across patients [30].

6.4. Vascular Imaging and Arterial Spin Labelling

6.4.1. Robust Cerebral Blood Flow Map Estimation in Arterial Spin Labeling

Participants: Camille Maumet, Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

Non-invasive measurement of Cerebral Blood Flow (CBF) is now feasible thanks to the introduction of Arterial Spin Labeling (ASL) Magnetic Resonance Imaging (MRI) techniques. To date, the low signal-to-noise ratio of ASL gives us no option but to repeat the acquisition in order to accumulate enough data to get a reliable signal. Perfusion signal is usually extracted by averaging across the repetitions. However, due to its zero breakdown point, the sample mean is very sensitive to outliers. A single outlier can thus have strong detrimental effects on the sample mean estimate. In this paper, we propose to estimate robust ASL CBF maps by means of M-estimators to overcome the deleterious effects of outliers. The behavior of this method is compared to z-score thresholding as recommended in [8]. validation on simulated and real data is provided. Quantitative validation is undertaken by measuring the correlation with the most widespread technique to measure perfusion with MRI: Dynamic Susceptibility weighted Contrast (DSC) [37].

6.4.2. A comprehensive framework for the detection of individual brain perfusion abnormalities using Arterial Spin Labeling

Participants: Camille Maumet, Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

Arterial Spin Labeling (ASL) enables measuring cerebral blood flow in MRI without injection of a contrast agent. Perfusion measured by ASL carries relevant information for patients suffering from pathologies associated with singular perfusion patterns. However, to date, individual identification of abnormal perfusion patterns in ASL usually relies on visual inspection or manual delineation of regions of interest. In this paper, we introduce a new framework to automatically outline patterns of abnormal perfusion in individual patients by means of an ASL template. We compare two models of normal perfusion and assess the quality of detections comparing an a contrario approach to the Generalized Linear Model (GLM) [33], [35], [36].

6.4.3. Using Negative Signal in Mono-TI Pulsed Arterial Spin Labeling to Outline Pathological Increases in Arterial Transit Times

Participants: Camille Maumet, Pierre Maurel, Jean-Christophe Ferré, Elise Bannier, Christian Barillot.

The presence of unexpected negative perfusion estimates has been sparsely discussed in the ASL literature. In the study of perfusion maps extracted from a single inversion time in ASL (mono-TI ASL), it is however common to deal with areas of significant negative signal. This is problematic since performing statistical analysis based on this data might therefore lead to inacurrate results. Though isolated negative values could be attributed to noise, clusters of significant negative signal should be explained by another phenomenon. Following [2], which outlined that negative values might arise due to increased transit times, we investigated this hypothesis based on real clinical datasets including healthy control and patient data [34].

6.4.4. An a contrario approach for the detection of activated brain areas in fMRI

Participants: Camille Maumet, Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

BOLD functional MRI (fMRI) is now a widespread imaging technique to study task-related activity in the brain. However, getting the areas of activation at the individual subject level is still an open issue. The standard massively univariate statistical analysis is usually performed after smoothing the data and makes use of a single p-value for final thresholding of the results. In group fMRI studies, the need for compensation of cross-subjects misregistrations clearly justifies the smoothing. However, at the individual level, where neat delineations of the activated areas are of interest, the use of gaussian smoothing as a pre-processing step is more questionable. In this paper, we propose to study the ability of an a contrario approach, recently adapted for basal perfusion abnormalities detection, to correctly detect areas of functional activity.

6.4.5. Compressive Matched Filter for Cerebral Blood Flow Quantification with ASL: sampling diversity or repetition?

Participants: Lei Yu, Pierre Maurel, Christian Barillot.

The Arterial Spin Labeling (ASL) is an MRI (Magnetic Resonance Imaging)-based perfusion technique which uses the magnetically tagged water as a freely diffusible tracer to measure perfusion non-invasively. This blood water is first labeled with a radio-frequency pulse in the neck of the patient. After a delay, called Inversion Time (TI), which allows the labeled blood to arrive in the brain, a labeled image of the brain is acquired. A control image is also acquired without labeling and the CBF (Cerebral Brain Flow) estimation is done on the difference between the control and labeled image. Classical method, Mono-TI, for CBF quantification is averaging repetitions with only one Inversion Time (TI) - the time delay between labeling and acquisition to allow the labeled blood to arrive the imaging slice. It improves the robustness to noise, however, cannot compensate the variety of Arterial Arrival Time (AAT).

In this work [45], Diverse-TI is proposed to exploit different TI sampling instants (sampling diversity) to improve the robustness to variety of AAT and simultaneously average repetitions with each TI (sampling repetitions) to improve the robustness to noise. Generally, the sampling diversity is relatively small and can be considered as compressed measurements, thus the Compressive Matched Filter (CMF) enlightened from sparsity is exploited to directly reconstruct CBF and AAT directly from compressed measurements. Meanwhile, regarding the CBF quantification performance, the compromise between the sampling repetition and sampling diversity is discussed and the empirical protocol to determine the sampling diversity is proposed.

The future works will consist in applying the parameter design protocol to guide the Diverse-TI technique in real ASL data acquisitions. Meanwhile, it is possible to extend CMF algorithm by considering additional priors to regularize the CBF estimation problem which might also improve the performance.

This work was done in collaboration with Remi Gribonval (Metiss team) [45].

6.4.6. Non-contrast enhanced neurovascular imaging

Participants: Elise Bannier, Hélène Raoult, Jean-Yves Gauvrit.

Detecting internal carotid artery (ICA) stenosis is a main challenge for the prevention of stroke, the third leading cause of death in the developed world. Novel non-contrast-enhanced MRA (NCE MRA) sequences have emerged as an alternative to traditional MRA approaches, especially for patients during pregnancy or with renal insufficiency.

Up to now, the inversion-prepared bSSFP NCE MRA approach has been applied to imaging of renal arteries or kidney transplants and only few studies focused on the ICA, using ECG-gating. The purpose of this first study was to assess the feasibility and image quality of an improved non-gated carotid NATIVE TrueFISP NCE MRA sequence providing an extended field of view as compared to Time-of-Flight (TOF) imaging. Sixteen healthy volunteers were included to evaluate different sequence parameter sets. In comparison to standard TOF, the used NCE MRA sequence offered equivalent to higher image quality along with larger coverage and shorter acquisition times. Improved image quality was achieved without ECG gating, which had been used in previous studies. A Partial Fourier scheme with an early acquisition of k-space center yielded higher image quality and signal intensity compared to a late acquisition.

A second study evaluated the non-contrast-enhanced ECG-gated 4D MRA combining arterial spin labeling (ASL) and bSSFP readout (bSSFP NCE 4D MRA) sequence to non invasively investigate morphological and hemodynamic patterns of cerebral arteriovenous malformations (AVM). Previous studies have shown high temporal resolution (50-100 ms), yet with temporal windows limited to a single cardiac cycle. This precludes the complete venous drainage analysis, which is necessary to evaluate AVM hemorrhagic risk. This study aimed at assessing the feasibility, quality and diagnosis performance of a bSSFP NCE 4D MRA sequence with a large acquisition time window over 2 cardiac cycles (2 RR) without a significant reduction of spatial resolution. Ten patients presenting AVM and referred to digital subtraction angiography (DSA) were included in the study. The 2-RR bSSFP NCE 4D MRA sequence yielded an image quality comparable to that of a corresponding 1-RR acquisition. AVM analysis, however, was improved due to a better depiction of venous drainage necessary to evaluate hemorrhagic risk. The simultaneous high-resolution morphologic and hemodynamic data also offered an especially accurate delineation of the nidus, target of the treatment.

6.4.7. ASLDEM : Arterial Spin Labeling At 3t In Semantic Dementia: Perfusion Abnormalities Detection And Comparison With Fdg-pet

Participants: Isabelle Corouge, Jean-Christophe Ferré, Elise Bannier, Christian Barillot, Jean-Yves Gauvrit.

Arterial Spin Labeling (ASL) is a non invasive perfusion imaging technique which has shown great diagnosis potential in dementia. However, it has never been applied to semantic dementia (SD), a rare subtype of frontotemporal lobar degeneration characterized by the gradual loss of conceptual knowledge, which is actually explored by a now well established marker of SD: ^{18}fluorodeoxyglucose-positron emission tomography (FDG-PET) imaging. Although ASL and FDG-PET respectively measure perfusion and metabolism, they have been shown to be strongly correlated. In this project, we explore the ability of ASL to detect perfusion abnormalities in SD in comparison with FDG-PET. We apply our analysis framework (implemented as part of the 'autoasl' and 'autoasltemplate' softwares) on patients and healthy subjects data from an ongoing clinical study with a focus on ASL data preprocessing and statistical analysis at the individual and group level. Preliminary results yield concordant observations between ASL and FDG-PET as well as expected hypoperfusions in SD, namely in the left temporal lobe, thus suggesting the potential of ASL to assess perfusion impairments in SD [29].

For this work, Aurore Esquevin was awarded the prize "Communication Jeune Chercheur 2012" at the "Journées Françaises de Radiologie (JFR)" conference.

6.5. Abnormal functional lateralization and activity of language brain areas in developmental dysphasia

6.5.1. Statistical analysis of white matter integrity for the clinical study of specific language impairment in children

Participants: Olivier Commowick, Aymeric Stamm, Camille Maumet, Jean-Christophe Ferré, Clément De Guibert, Christian Barillot.

Children affected by Specific Language Impairment (SLI) fail to develop a normal language capability. To date, the etiology of SLI remains largely unknown. It induces difficulties with oral language which cannot be directly attributed to intellectual deficit or other developmental delay. Whereas previous studies on SLI focused on the psychological and genetic aspects of the pathology, few imaging studies investigated defaults in neuroanatomy or brain function. We propose to investigate the integrity of white matter in Specific Language Impairment thanks to diffusion Magnetic Resonance Imaging. An exploratory analysis was performed without a priori on the impaired regions. A region of interest statistical analysis was performed based, first, on regions defined from Catani's atlas and, then, on tractography-based regions. Both the mean fractional anisotropy and mean apparent diffusion coefficient were compared across groups. To the best of our knowledge, this is the first study focusing on white matter integrity in specific language impairment. 22 children with SLI and 19 typically developing children were involved in this study. Overall, the tractography-based approach to group comparison was more sensitive than the classical ROI-based approach. Group differences between controls and SLI patients included decreases in FA in both the perisylvian and ventral pathways of language, comforting findings from previous functional studies. This work was performed in collaboration with Emmanuel Vallée, Clément de Guibert, Catherine Allaire and Elisabeth Le Rumeur.

ACES Project-Team

4. New Results

4.1. Spatial Computing approach and RFIDs

Participants: Michel Banâtre, Paul Couderc [contact], Yann Glouche, Arnab Sinha.

In the line of our previous research in pervasive computing, we are working on spatial computing approaches in the context of RFID. Spatial computing consists in data structures and computing processes directly supported by physical objects. RFID is an attractive technology for supporting spatial computing, enabling any object to interact in a smart environment. Traditionnal RFID solutions use a logical model, where the RFID tags are simple identifiers referring to data in a remote information system. In our approach, we use the memory of the tags to build self-contained data structures and self-describing objects. While featuring interesting properties, such as autonomous operation and high scalability, this approach also raises difficult challenges: the memory capacity of the tags is very limited, requiring compact and efficient data structures.

Our research in the context of domestic waste management is broadly investigating the use of RFID at item level to provide early waste sorting, to avoid incompatible mix of waste and to prevent hazards [3], [4]. Several innovative aspected are studied in this project. First, the design of an autonomous computing architecture for the waste items and smart containers, enabling early processing in the waste management: for example waste bags can be accepted or rejected accordingly to their content and its conformance with the recipient container. Hazard prevention and human operator safety can also be improved with the knowledge of the nature of the waste.

Autonomy is important as it would be possible to depend on a remote information system for each waste insertion, due to obvious scalability, energy and network costs. An ontology based system has been proposed to determine the possible interactions of tagged products based on their properties and the external conditions [6]. This ontological model is simple enough to be supported entirely by a low power embedded computer at the container level, but can still support the waste application requirements. An unconventional aspect in this architecture is that semantic properties are directly written in the RFID tags, instead of semantic-less identifiers typically used in most RFID applications.

A second innovative aspect of the research is to consider the set of containers in a city as a particular case of sensor network, and developing energy efficient protocol to enable information reporting to a supervising infrastructure.

In the context of this research, some limitations of existing RFID technology become challenging: unlike standard RFID application scenarios, pervasive computing often involves uncontrolled environment for RFID, where tags and reader have to operate in much more difficult situations that those usually encountered or expected for classical RFID systems. In a near future, we seek to work with a team who has a strong expertise in antenna design and radio signal behaviour.

4.2. Integrity checking with coupled objects

Participants: Michel Banâtre [contact], Paul Couderc, Jean-Francois Verdonck.

While the computing and telecommunication worlds commonly use digital integrity checking, many activities from the real world do not benefit from automatic integrity control mechanisms. RFID technology offers promising perspectives for facing this problem, but also raises strong privacy concerns as most of the RFID-based systems rely on global identification and tracking. In 2011, we have designed Ubi-Check to provide an approach aiming at coupling physical objects and enabling integrity control built on local interactions, without the support of a global information system. Ubi-Check led to the development of various novel applications running quite on the same technology. But the possibility of defining hierarchical couplings was lacking.

This is why we have studied and and designed the Ubi-Tree environment in 2012, which strives to deal with those new requirements. Ubi-Tree relies on a structure in which physical objects (also called fragments) are seen as external nodes of a tree that we call coupling tree. External nodes of a tree are called leaves. In the system, internal nodes are called coupling nodes. Each fragment embeds an RFID tag supporting coupling data. Coupling data stores the coupling tree. Each internal node can be checked, which means a lacking, illegally forged or corrupted node can be detected at any depth of a coupling.

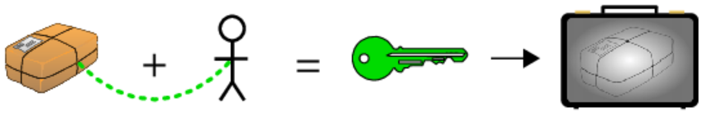


Figure 2. Key to a Ubi-Post briefcase

The Ubi-Tree environment has been experimented through a content-oriented security solution for high value shipping: the Ubi-Post briefcase. Sending sensitive documents or parcels over a delivery service can be a hazardous operation. Goods can be picked up by a fake courier, genuine items can be swapped with copies, the parcel may be received or opened by someone else than the supposed recipient and some items can be missing at the delivery time. As some very high value items are sent over such services, security is critical. We proposed the Ubi-Post briefcase system, a pervasive content-oriented security solution for high value shipping based on the Ubi-Tree physical object coupling software and RFID equipment. The aim of a shipping service is to provide transportation of goods from a sender to the recipient, so the system must ensure that the coupling would be handed over to the recipient. For that purpose, coupled tags will carry an identifier corresponding to the recipient as additional data. Then, the only way to unlock a Ubi-Post briefcase is to insert a recipient card which tag ID is the one expected by the coupling (see figure 2). The Ubi-Post briefcase embeds the same equipment as the coupling station, plus a battery, an HF near field card reader, and a locking mechanism (see Figure 3).

We have produced an interface for users to be sure that the association between RFID tag and physical object is the one that is perceived by our coupling software. The key idea was to be able to identify in the right way the RFID tag associated to a physical object when we place one physical object onto the support of the antenna linked to the RFID reader. The position of this object, and the tag associated to this object, in the physical space is determined using a camera coupled with an image recognition algorithm. The result is displayed onto a touch screen. In that way, when we want to couple a set of physical objects, we place sequentially all these objects onto the support of the antenna, and from the image of these objects displayed onto the touch screen we touch those we want to couple and activate the coupling operation. This solution is now fully functional.

4.3. Pervasive support for Smart Homes

Participants: Michele Dominici, Bastien Pietropaoli, Sylvain Roche, Frédéric Weis [contact].

A smart home is a residence equipped with information-and-communication-technology (ICT) devices conceived to collaborate in order to anticipate and respond to the needs of the occupants, working to promote their comfort, convenience, security and entertainment while preserving their natural interaction with the environment.

The idea of using the Ubiquitous Computing paradigm in the smart home domain is not new. However, the state-of-the-art solutions only partially adhere to its principles. Often the adopted approach consists in a heavy deployment of sensor nodes, which continuously send a lot of data to a central elaboration unit, in charge of the difficult task of extrapolating meaningful information using complex techniques. This is a

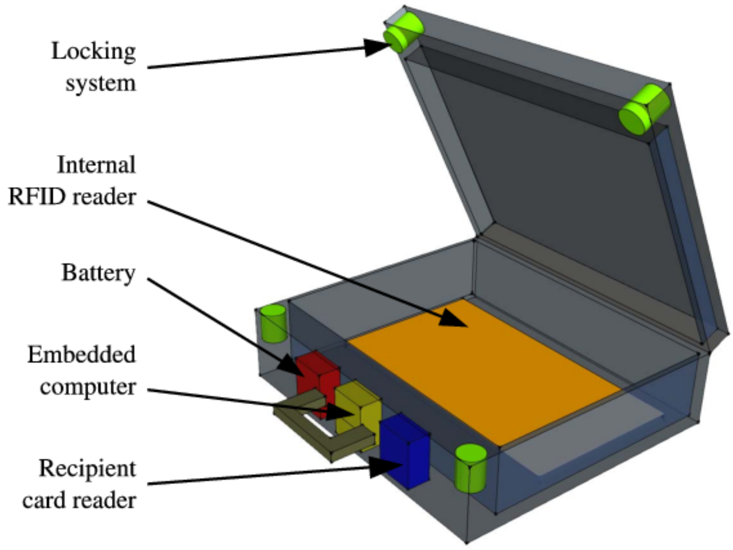


Figure 3. 3D view from the internal components of the Ubi-Post briefcase

logical approach. ACES proposed instead the adoption of a *physical approach*, in which the information is spread in the environment, carried by the entities themselves, and the elaboration is directly executed by these entities "inside" the physical space. This allows performing meaningful exchanges of data that will thereafter need a less complicate processing compared to the current solutions. The result is a smart home that can, in an easier and better way, integrate the context in its functioning and thus seamlessly deliver more useful and effective user services. Our contribution aims at implementing the physical approach in a domestic environment, showing a solution for improving both comfort and energy savings.

Most existing smart home solutions were designed with a technology-driven approach. That is, the designers explored which services, functionalities, actions and controls could be performed exploiting available technologies. This led to solutions for human activity recognition relying on wearable sensors, microphones or video cameras. Those technologies may be difficult to deploy and get accepted in real-world households, because of convenience and privacy concerns. Many people have concerns on carrying equipments or feeling observed or recorded while living their private life. This could seriously impact the acceptability of the smart home system or reduce its diffusion in real households. To avoid such kind of issues, we designed our system with an acceptability-driven approach. That is, we selected technologies that respond to the constraints of a real-world deployment of the future smart home system, namely, convenience and privacy concerns. We decided to take a very conservative approach, choosing technologies that are as unobtrusive as possible, in order to explore the frontiers of what can be done in a smart home with a very limited instrumentation. Following the same considerations, the adopted technologies and techniques had to guarantee a fast and easy configuration, ultimately allowing a plug-and-play deployment.

4.3.1. Design and implementation of a system architecture

In 2012, we have designed and experimented a system architecture of a smart home prototype currently under development. It is the demonstrator of an interdisciplinary project that brings together industrials and researchers, from the fields of ubiquitous computing and cognitive ergonomics. The aim is to develop a smart home system that is able to prevent energy waste and preserve inhabitants' comfort. The key requirement is to provide functionalities that are seamlessly adapted to ongoing situations and activities of inhabitants, avoiding bothering them with inappropriate interventions. The architecture of such a system has been designed so as to respect the principles and constraints illustrated in the introduction of this section. Namely, we have chosen the necessary equipments among those that should guarantee privacy preservation and high acceptability. When designing the algorithms for context and situation recognition and the human-computer interaction aspects of the system, we have kept in mind the model of human activity described in the previous section. Finally, we have designed the architecture of the system so as to realize successive abstraction of contextual information and to allow uncertainty, imprecision and ignorance to flow between the layers [2].

4.3.2. Layered architecture

The system architecture relies on the principles of the ubiquitous computing paradigm. It also draws its inspiration from the work of Coutaz, who suggest a four-layer model to build context-aware applications. The first layer, "sensing", is in charge of sensing the environment. It is realized by augmented appliances and physical sensors. The augmented household appliances provide information about their state, while the sensors measure physical phenomena (sound level, motion, vibration, etc.). The second layer, called "perception", realizes the abstraction from the raw data. These are processed to obtain more abstract information about the context (e.g. the detection of presence in a room can be obtained combining motion, sound and vibration measures). "Situation and context identification", the third layer, identifies the occurring situations and the activities of inhabitants. For instance, the fact that a given moment a person is ironing can be modeled combining the information that a person is present in a room with the fact that the iron is on and that it is being moved. The top layer, called "exploitation", provides contextual information to applications. More specifically, the contextual information is used to adapt the behavior of the augmented appliances in a semi-automatic way and to allow lowly interruptive takeover by inhabitants.

4.3.3. Design and experimentation of the "perception" layer

In the second layer called "perception", raw sensor data are processed to obtain more abstract information about context called Context Attributes. These are small pieces of context easily understandable by humans and that can be provided to the upper layer. Examples of Context Attributes are the presence, the number of people in a room or the posture of someone. Some raw data are immediately exploitable, like temperature or light level. Others require data fusion in order to obtain more abstract contextual information, such as inhabitants' presence or movement. A certain number of sensors is necessary to obtain sufficient certainty when fusing data, as redundancy can significantly increase the reliability of the sources. Furthermore, heterogeneous sensors allow collecting different physical measurements that can enrich the data fusion process.

Data fusion is a large problem. Many theories offer tools to handle it. In our approach, the main aim of the perception layer is to abstract imperfect raw data to make it computable by higher level reasoning algorithms. Data may be imperfect for different reasons:

- Randomness, due to physical systems (in our case, sensors).
- Inconsistency, due to overload of data or conflicting sources.
- Incompleteness, due to loss of data which may easily happen with wireless communication.
- Ambiguity (or fuzziness), due to models or to natural language imprecision.
- Uncertainty, due to not fully reliable sources.
- Bias, due to systematic errors.
- Redundancy, due to multiple sources measuring the same parameter.

In order to manage many of those imperfections and respect the theoretical constraints, we decided to use as a first layer of abstraction the belief functions theory (BFT). The BFT can be seen as a generalization of the Bayesian theory of subjective probability. It can be used to model probabilities if only atomic focal sets are used in mass functions. Thus, it is totally possible to mix probabilities with real belief functions.

In our approach, we considered that sensors should duce belief for a certain amount of time after the measures because of the continuity of studied context. For instance, a motion sensor in a room could be able to induce a belief on the presence of someone for a longer time than the exact moment at which the measure has been obtained. It is a matter of physical system with inertia. In this example, it is easy to take into account that physical persons cannot move too fast and thus will certainly be there for some seconds before they can exit the room. Thus, this little example brings two questions: how to build evidence from raw data and how to take into account evidence over time? We proposed a simple method already existing to build belief functions from raw data and propose an improvement to take into account timed evidence [5].

4.3.4. Design and experimentation of the "situation and context identification" layer

"Situation and context identification", the third layer, identifies the occurring situations and the activities of inhabitants. For instance, the fact that a given moment a person is ironing can be modeled combining the information that a person is present in a room with the fact that the iron is on and that it is being moved. Having obtained the Context Attributes through abstraction from the raw sensor data, the system has to reason about context, in order to infer higher-level context information, needed to make decisions concerning the functionalities to offer to inhabitants. We needed a unified theory for modeling contextual information, also offering a generic framework for applying different reasoning techniques to infer higher-level context.

We adopted a situation-centric modeling and reasoning approach called *Context Spaces*, based on a unified context modeling and reasoning theory. Using this theory, interesting situations can be modeled as combinations of basic contextual information provided both by a sensor-data-fusion technique and by augmented appliances. Adapted functionalities can be provided when the interesting situations are triggered. The recognition of ongoing situations is made possible by reasoning about available context information. The Context Spaces theory allows managing and propagating uncertainty and ignorance, reasoning on ambiguous contexts and assessing the degree of uncertainty of the resulting inference. It also provides tools to reason on complex logical expressions that combine elementary situations. The use and the extension of the Context Spaces is the core of a PhD thesis that has been finished at the end of 2012 by Michele Dominici (to be defended in March 2013).

4.3.5. Uncertainty and ignorance management

Given the gap between contextual capture capabilities of our architecture and actual complexity of real-world human activities and context, an important issue arises: the management of uncertainty and ignorance. If contextual information has to be abstracted in successive steps, sources are not always reliable. In particular, uncertainty is intrinsic to the physical sensors that are used in the capture. Thus, the uncertainty of lower abstraction layers will negatively impact the inference and decisions of the upper layers. Furthermore, due to the contextual gap illustrated above, any computing model that tries to represent the complexity of real activity will be affected by a certain degree of uncertainty. This reflects on the recognition of the activity itself and can lead to wrong conclusions, which in turn negatively impact the provision of adapted functionalities to inhabitants. As a consequence, we considered that information about uncertainty and ignorance has to be propagated, cumulated and considered at every layer of our pervasive architecture. Whenever the level of uncertainty becomes excessively high, the system tried to evaluate the tradeoff between the potential benefit of providing the right functionality and the risk associated with an unsuitable functionality, which would be provided in case the situation has not been correctly recognized.

ADAM Project-Team

6. New Results

6.1. Software Product Lines

In terms of Software Product Lines [92], we work in four different directions. First, we define a SPL framework for Cloud Computing called SALOON [62] to face challenges in terms of application configuration, cloud platform configuration [59] and deployment automation [58]. Second, we use Dynamic Software Product Lines (DSLP) for mobile applications [21], in order to support self-adaptation of context-aware applications in ubiquitous environments [56] and Wireless Sensor Networks (WSNs) [36]. In both cases, Constraint Satisfaction Problem (CSP) techniques are used in order to find a suitable configuration for the current environment state and to deal with contradictory dimensions (e.g., accuracy and energy saving) in the decision making process. Third, in the YourCast project [76], we work in a Composite SPL for Broadcasting System by identifying the main issues that we need to deal with when defining such kind of SPL. Finally, we define an operator to compute syntactic and semantic differences between feature models [24].

6.2. Software Evolution

The adaptive software paradigm supports the definition of software systems that are continuously adapted at run-time. An adaptation activates multiple features in the system, according to the current execution context (*e.g.*, CPU consumption, available bandwidth). However, the underlying approaches used to implement adaptation are ordered, *i.e.*, the order in which a set of features are turned on or off matters. Assuming feature definition as etched in stone, the identification of the right sequence is a difficult and time-consuming problem. We propose here a composition operator that intrinsically supports the commutativity of adaptations [50]. Using this operator, one can minimize the number of ordered compositions in a system. It relies on an action-based approach, as this representation can support preexisting composition operators as well as our contribution in an uniform way. This approach is validated on the Service-Oriented Architecture domain, and is implemented using first-order logic.

6.3. Green Middleware

The energy consumption of ICT is widely acknowledged as continuously growing over years and its carbon footprint can now be compared to the avionics domain. While green computing has emerged as a new research area concerned with the optimization of the energy consumption of large-scale systems, such as datacenters, our project-team investigates the analysis of the energy consumption from a software engineering point of view. In particular, we developed e-Surgeon, a middleware framework to estimate the power consumption of legacy software at various levels of granularity. With respect to this objective, the first result we obtained in [52] relates to an evaluation of the impact of programming languages and programming styles on the energy consumption of applications. While the current trend in application servers is to adopt interpreted languages (*e.g.*, JavaScript, Python) on the server side, our preliminary results highlight that these languages impose a large overhead to the energy consumption of the resulting system. In [53], this preliminary result is further investigated by identifying energy hotspots within legacy application servers. To do so, we automatically instrument the code of the application server to analyze how the energy is consumed by the application server under various stress scenarios. Our results show that the energy is mostly consumed by a restricted number of classes and methods of these application servers, thus giving hints to software developers on candidate snippets for optimization.

6.4. Human-Competitive Software Engineering

Frequently asked questions (FAQs) are a popular way to document software development knowledge. As creating such documents is expensive, we have invented an approach for automatically extracting FAQs from sources of software development discussion, such as mailing lists and Internet forums, by combining techniques of text mining and natural language processing. We applied the approach to popular mailing lists and carried out a survey among software developers to show that it is able to extract high-quality FAQs that may be further improved by experts. This research has been published at the International Conference on Software Engineering (ICSE'2012 [40]), the flagship conference in the domain. This work takes place in our line of research on "human competitive software engineering", where we try to replace manual tasks requiring costly human skills (such as documentation writing or bug fixing) by automated or semi-automated approaches.

6.5. Reconfigurable Middleware

In the context of our collaboration with the Thales company, especially via the PhD of Jonathan Labéjof defended on 20 December 2012, we obtained some results in the domain of reconfigurable message-oriented middleware (MOM). MOM are a particular class of middleware systems that promote asynchronous communications and weak coupling between communicating entities. They are of particular interest for the design of Systems of Systems (SoS). In this context, we worked on a method for reconfiguring quality of service properties in MOM. The idea is to be able change the properties of communication channels without stopping these channels. We obtained this by defining a bijection between the characteristics of these channels and a component-based software architecture for which we already have means of reconfiguration with our previous results on the FRASCATI platform (see Section 5.5). By this way, reconfiguring the quality of service of a channel is akin to reconfiguring its associated component-based software architecture. This result has been applied to MOM platforms that conform to the OMG DDS standard.

This result has been the topic of a patent application [106] that was filled in Europe in July 2011 and in the US in July 2012. The results were also presented in the SCDI workshop at the EDOC 2012 conference [44].

ALGORILLE Project-Team

6. New Results

6.1. Structuring applications for scalability

In this domain we have been active on several research subjects: efficient locking interfaces, data management, asynchronism, algorithms for large scale discrete structures and the use of accelerators, namely GPU.

In addition to these direct contributions within our own domain, numerous collaborations have permitted us to test our algorithmic ideas in connection with academics of different application domains and through our association with SUPÉLEC with some industrial partners: physics and geology, biology and medicine, machine learning or finance.

6.1.1. Efficient linear algebra on accelerators.

Graphics Processing Units have evolved to fully programmable parallel vector-processor sub-systems. We have designed several parallel algorithms on GPUs, and integrated that level of parallelism into larger applications including several other levels of parallelism (multi-core, multi-node,...). In this context, we also have studied energy issues and designed some energy performance models for GPU clusters, in order to model and predict energy consumption of GPU clusters.

The PhD thesis of Wilfried Kirschenmann, has been a collaboration with EDF R&D and was co-supervised by S. Vialle and Laurent Plagne (EDF SINETICS). It has given rise to a DSEL based on C++ and to a unified generic library that adapts to multi-core CPUs, multi-core CPUs with vector units (SSE or AVX), and GPUs. This framework allows to implement linear algebra operations originating from neutronic computations, see [22].

The PhD thesis of Thomas Jost, co-supervised by S. Contassot-Vivier and Bruno Lévy (Alice INRIA team) deals with specific algorithms for GPUs, in particular linear solvers. He has also worked on the use of GPUs within clusters of workstations via the study of a solver of non-linear problems [17]. The defense of this thesis is planned in January 2013.

6.1.2. Combining locking and data management interfaces.

Handling data consistency in parallel and distributed settings is a challenging task, in particular if we want to allow for an easy to handle asynchronism between tasks. Our publication [4] shows how to produce deadlock-free iterative programs that implement strong overlapping between communication, IO and computation; [21] extends distributed lock mechanisms and combines them with implicit data management.

A new implementation (ORWL) of our ideas of combining control and data management in C has been undertaken, see 5.5 . A first work has demonstrated its efficiency for a benchmark application [18]. Our current efforts concentrate on the implementation of a complete application (an American Option Pricer) that was chosen because it presents a non-trivial data transfer and control between different compute nodes and their GPU. ORWL is now able to handle such an application seamlessly and efficiently, a real alternative to home made interactions between MPI and CUDA.

6.1.3. Discrete and continuous dynamical systems.

The continuous aspect of dynamical systems has been intensively studied through the development of asynchronous algorithms for solving PDE problems. We have focused our studies on the interest of GPUs in asynchronous algorithms [17]. Also, we investigate the possibility to insert periodic synchronous iterations inside the asynchronous scheme in order to improve the convergence detection delay. This is especially interesting on small/middle sized clusters with efficient networks. Finally, we investigate other optimizations like load balancing. For this last subject, the SimGrid environment has revealed itself to be a precious tool to perform feasibility tests and benchmarks for this kind of algorithms on large scale systems. It has been successfully used to evaluate an asynchronous load balancing algorithm [37].

In 2011, the PhD thesis of Marion Guthmuller, supervised by M. Quinson and S. Contassot-Vivier, has started on the subject of model-checking distributed applications inside the SimGrid simulator [20]. This is also the opportunity of designing new tools to study more precisely the dynamics of discrete or continuous systems. See the simulation part in Section 6.3.2 for more details on this PhD.

6.2. Transparent resource management

6.2.1. Client-side cloud broker.

Integrating the 'pay-as-you-go' pricing model commonly used in IaaS clouds is an important question which profoundly changes the assumptions for job scheduling. From the observation that in most commercial solutions the price of a CPU cycle is identical, be the CPU a fast or slow one, several schedulings may be derived for a same price but with different makespans. Hence, in a context where resources can be started on-demand, scheduling strategies must include a decision process regarding the scaling (number of resources used) of the platform and the types of resources rented over time. In [24], we have studied the impact of these two factors on classic job scheduling strategies applied to bag-of-tasks workloads. The results show that shorter makespans can be achieved through scaling at no extra cost, while using quicker CPUs largely increases the price of the computations. More importantly, we show the difficulty to predict the outcomes of such decisions, which requires to design new provisioning approaches.

6.3. Experimental Methodologies

6.3.1. Overall improvement of SimGrid

2012 was the last year of the USS-SimGrid project granted by the ANR. We thus capitalized the results of the first project by properly releasing them in the public releases. Parallel simulation is now stable enough to be used in practice by our users. In addition, the framework is now able to simulate millions of processes without any particular settings in C. The java bindings were also improved to simulate several hundred thousand processes out of the box [25].

This year was also the first year of the SONGS project, also funded by the ANR. This project is much larger that the previous one, both in funding and targets. In surface, SONGS aims at increasing the scope of the SimGrid simulation framework by enabling the Cloud and HPC scenarios in addition to the existing Grid and P2P ones. Under the hood, it aims at providing new models specifically designed for these use cases, and also provide the necessary internal hooks so that users can modify the used models by themselves.

This project is well started, with three plenary meetings and a user conference organized over the year, but no new publication resulted of this work yet. The first work toward increasing the simulation versatility, initiated last year, was published this year[14]

6.3.2. Dynamic verification of liveness properties in SimGrid

A full featured model-checker is integrated to SimGrid since a few years, but it was limited to the verification of safety properties. We worked toward the verification of liveness properties in this framework. The key challenge is to quantify the state equality at state level, adding and leveraging introspection abilities to arbitrary C programs.

This constitutes the core of the PhD thesis of M. Guthmuller, started last year. A working prototype was developed during this year, described in an initial publication [20].

6.3.3. Grid'5000 and related projects

We continued to play a key role in the Grid'5000 testbed in 2012. Lucas Nussbaum, being delegated by the executive committee to follow the work of the technical team, was heavily involved in the recent evolutions of the testbed (network weathermaps, storage management, etc.) and in other activities such as the preparation of the Grid'5000 winter school. We were also involved in a publication [33] which is a follow-up to the workshop on *Supporting Experimental Computer Science* held during SC'11, and in another publication [32] describing the recent advances on the Grid'5000 testbed in order to support experiments involving virtualization at large scale.

More specifically, our involvement in the *OpenCloudWare* project led us to design several tools that ease the deployment of Cloud stacks on Grid'5000 for experimental purposes. Those tools were also used during an internship that was co-advised with the *Harmonic Pharma* start-up, exploring how complex bio-informatics workflows could be ported to the Cloud.

On the institutional side, we will also play a central role in the *Groupement d'Intérêt Scientifique* that is currently being set up, since Lucas Nussbaum is a member of both the *bureau* and of the *comité d'architectes*.

6.3.4. Distem – DISTributed systems EMulator

In the context of ADT Solfége, we continued our work on Distem. Three releases were made over the year, with several improvements and bug fixes, including support for variable CPU and network emulation parameters during an experiment. See http://distem.gforge.inria.fr/ for more information, or our paper accepted at PDP'2013 [26].

6.3.5. Kadeploy3 – scalable cluster deployment solution

Thanks to the support of ADT Kadeploy3, many efforts were carried out on Kadeploy3. Two releases were made, including many new features (many improvements to the handling of parallel commands and to the inner automaton for more fault-tolerant deployments; use of Kexec for faster deployments) as well as bug fixes.

Kadeploy3 was featured during several events (*journée 2RCE*, *SuperComputing 2012*), and in two publications: one unsuccessfully submitted to LISA'2012 [35], one accepted in USENIX *;login:* [13].

Finally, Kadeploy3 was also the basis of submissions to the *SCALE challenge held with CCGrid'2012*, of which we were finalists, and of the winner challenge entry at *Grid'5000 winter school 2012*.

6.3.6. Business workflows for the description and control of experiments

We are exploring the use of Business Process Modelling and Management for the description and the control of complex experiments. In [28], we outlined the required features for an experiment control framework, and described how business workflows could be used to address this issue. In [27] and [15], we described our early implementation of XPFlow, a experiment control engine relying on business workflows paradigms.

6.3.7. Towards Open Science for distributed systems research

One of our long term goal on experimental methodologies would be the advance of an Open Science in the research domain of Distributed Systems. Scientific tools would be sufficiently assessed and easily combined when necessary, and scientific experiments would be perfectly reproducible. These objectives are still very ambitious for the researches targeting distributed systems.

In order to precisely evaluate the path remaining toward these goals, and try addressing some of the challenges that they pose, we currently host Maximiliano Geier as an Inria intern. While most researchers try to answer brilliant scientific questions with simple scientific methodologies, he is asked to answer a simple question (on the adaptation of the BitTorrent protocol to high bandwidth networks) using an advanced scientific methodology. We are also surveying the experimental methodology used in top tier conferences to gain further insight on this topic.

In addition, we are organizing Realis, an event aiming at testing the experimental reproducibility of papers submitted to Compas'2013. Associated to the Compas'13 conference, this workshop aims at providing a place to discuss the reproducibility of the experiments underlying the publications submitted to the main conference. We hope that this kind of venue will motivate the researchers to further detail their experimental methodology, ultimately allowing others to reproduce their experiments.

ARLES Project-Team

6. New Results

6.1. Introduction

The ARLES project-team investigates solutions in the forms of languages, methods, tools and supporting middleware to assist the development of distributed software systems, with a special emphasis on mobile distributed systems enabling the ambient intelligence/pervasive computing vision. Our research activities in 2012 have focused on the following areas:

- Dynamic interoperability among networked systems toward making them eternal, by way of on-the-fly generation of connectors based on adequate system models (§ 6.2);
- Pervasive service-oriented software engineering, focusing on supporting service composition in an increasingly heterogeneous and dynamic networking environment, while enforcing quality of service (§ 6.3);
- Service oriented middleware for the ultra large scale future Internet of Things (§ 6.4);
- Abstractions for enabling domain experts to easily compose applications on the Internet of Things (§ 6.5); and
- The use of Requirement Engineering techniques for enabling systems to be self-adaptive under uncertainty (§ 6.6).

6.2. Emergent Middleware Supporting Interoperability in Extreme Distributed Systems

Participants: Emil Andriescu, Amel Bennaceur, Luca Cavallaro, Valérie Issarny, Daniel Sykes.

Interoperability is a fundamental challenge for today's extreme distributed systems. Indeed, the high-level of heterogeneity in both the application layer and the underlying infrastructure, together with the conflicting assumptions that each system makes about its execution environment hinder the successful interoperation of independently developed systems. A wide range of approaches have thus been proposed to address the interoperability challenge. However, solutions that require performing changes to the systems are usually not feasible since the systems to be integrated may be legacy systems, COTS (Commercial Off-The-Shelf) components or built by third parties; neither are the approaches that prune the behavior leading to mismatches since they also restrict the systems' functionality. Therefore, many solutions that aggregate the disparate systems in a non-intrusive way have been investigated. These solutions use intermediary software entities, called *mediators*, to interconnect systems despite disparities in their data and/or interaction models by performing the necessary coordination and translations while keeping them loosely-coupled. However, creating mediators requires a substantial development effort and a thorough knowledge of the application-domain, which is best understood by domain experts. Moreover, the increasing complexity of today's distributed systems, sometimes referred to as Systems of Systems, makes it almost impossible to develop 'correct' mediators manually. Therefore, formal approaches are used to synthesize mediators automatically.

In light of the above, we have introduced the notion of *emergent middleware* for realizing mediators. Our research on enabling emergent mediators is done in collaboration with our partners of the CONNECT project (§ 7.2.1.1). Our work during the year has more specifically focused on:

- **Architecture enabling emergent middleware.** We have been finalizing, together with our partners in the CONNECT project, the definition of an overall distributed system architecture supporting emergent middleware, from the discovery of networked systems to the learning of their respective behavior and synthesis of emergent middleware enabling them to interoperate [31].

- **Affordance inference.** We have proposed an ontology-based formal model of networked systems based on their affordances (high-level functionalities), interfaces, behavior, and non-functional properties, each of which describes a different facet of the system in a way similar to the service description promoted for semantic Web services. However, legacy systems do not necessarily specify all of the aforementioned facets. Therefore, we have explored techniques to infer the affordance by using textual descriptions of the interface of networked systems. More specifically, we rely on machine learning techniques to automate the inference of the affordance from the interface description by classifying the natural-language text according to a predefined ontology of affordances. In a complementary way, CONNECT partners investigate protocol-learning algorithms to learn the behavior of networked systems on the fly [17].
- **Mediator synthesis for emergent middleware.** We focus on systems that have compatible functionalities, i.e., semantically matching affordances, but are unable to interact successfully due to mismatching interfaces or behaviors. To solve such mismatches, we propose a *mapping-based* approach, whose goal is to automatically synthesize a mediator model that ensures the *safe* interaction of functionnally compatible systems, i.e., deadlock-freedom and the absence of unspecified receptions. Our approach combines semantic reasoning and constraint programming to identify the semantic correspondence between networked systems' interfaces, i.e., *interface mapping*. Unlike existing approaches that only tackle the one-to-one correspondence between actions and for which we investigated a solution using ontology-based model checking [16], the proposed mapping-based approach handles the more general cases of one-to-many and many-to-many mappings. This work has resulted in a supporting software prototype that allows validating the approach; related publication is under writing. A further key research issue we are addressing in emergent middleware is the study of cross-paradigm interaction so as to enable interoperability among highly heterogeneous services (e.g., an IT-based service will likely interact using the client-service scheme while thing-based services rather adopt asynchronous protocols). Toward that goal, we are studying abstract models associated with popular interaction paradigms and higher level, generic interaction paradigms to define cross-paradigm mappings that respect the behavioral semantics of the interacting systems.
- **Automated mediation for cross-layer protocol interoperability.** While existing approaches to interoperability consider either application or middleware heterogeneity separately, we believe that in real world scenarios this does not suffice: application and middleware boundaries are ill-defined and solutions to interoperability must consider them in conjunction. As part of our recent work, we have proposed such a solution, which solves cross-layer interoperability by automatically generating parsers and composers that abstract physical message encapsulation layers into logical protocol layers, thus supporting application layer mediation. Specifically, we support the automated synthesis of mediators at the application layer using the mapping-based approach discussed above, while we introduce *Composite Cross- Layer (CCL) parsers and composers* to handle cross-layer heterogeneity and to provide an abstract representation of the application data exchanged by the interoperating components. In particular, we associate the data embedded in messages with annotations that refers to concepts in a domain ontology. As a result, we are able to reason about the semantics of messages in terms of the operations and the data they require from or provide to one another and automatically synthesize, whenever possible, the appropriate mediators. We have demonstrated the validity of our approach by using the framework to solve cross-layer interoperability between existing conference management systems.
- **Models@run.time.** We have recently integrated the notion of *Models@run.time* in our research towards emergent middleware. We use Models@run.time to extend the applicability of models and abstractions to the runtime environment. As is the case for software development models, a runtime model is often created to support reasoning. However, in contrast to development models, run-time models are used to reason about the operating environment and runtime behavior, and thus these models must capture abstractions of runtime phenomena. Different dimensions need to be balanced, including resource-efficiency (time, memory, energy), context-dependency (time, location, platform), as well as personalization (quality-of-service specifications, profiles). The hypothesis is

that because Models@run.time provide meta-information for these dimensions during execution, run-time decisions can be facilitated and better automated.

Thus, we anticipate that Models@run.time will play an integral role in the management of extremely distributed systems. Our way of using runtime models captures syntax and also semantics of behaviour and supports runtime reasoning. Prior models@run.time approaches have generally concentrated on architectural-based runtime models and self-adaptation of existing software artifacts. However, such artefacts cannot always be produced in advance, and we believe that models@runtime have a fundamental role to play in the production of dynamic, adaptive, and on-the-fly software as investigated in the context of emergent middleware [8]. Specifically, two important methods underpin our approach: *i*) automatic inference of the required runtime models during execution and their refinement by exploiting learning and synthesis techniques; and *ii*) using these models for a dynamic software synthesis approach, where mediators are formally characterized (using LTS) to allow the runtime synthesis of software.

In order to enable emergent middleware, we have shown how systems can infer information to build runtime models during execution. Importantly, ontologies were exploited to enrich the runtime models and facilitated the mutual understanding required to perform the matching and mapping between the networked heterogeneous systems. Such reasoning about information that was not necessarily known before execution, is in contrast to the traditional use of models@run.time.

6.3. Revisiting the Abstractions of Service Oriented Computing for the Future Internet

Participants: Dionysis Athanasopoulos, Sandrine Beauche, George Bouloukakis, Oleg Davidyuk, Nikolaos Georgantas, Valérie Issarny, Ajay Kattepur.

A software architecture style characterizes, via a set of abstractions, the types of: components (i.e., units of computation or data stores), connectors (i.e., interaction protocols) and possibly configurations (i.e., system structures) that serve to build a given class of systems. As such, the definition of a software architectural style is central toward eliciting appropriate design, development and runtime support for any family of systems. The service oriented architecture style may then be briefly defined as follows: (1) components map to services, which may be refined into consumer, producer or prosumer services; (2) connectors map to traditional client-service interaction protocols; and (3) configurations map to compositions of services through (service-oriented) connectors, e.g., choreography and orchestration structures. While the service-oriented architecture style is well suited to support the development of Internet-based distributed systems, it is largely challenged by the Future Internet that poses new demands in terms of sustaining *ities* such as scalability, heterogeneity, mobility, awareness & adaptability that come in extreme degrees compared to the current Internet. Therefore, we have been working on eliciting software architectural abstractions for the Future Internet by building upon the service-oriented architecture style, as well as on applying them to system design, development and execution.

Complex distributed applications in the Future Internet will be to a large extent based on the open integration of extremely heterogeneous systems, such as lightweight embedded systems (e.g., sensors, actuators and networks of them), mobile systems (e.g., smartphone applications), and resource-rich IT systems (e.g., systems hosted on enterprise servers and Cloud infrastructures). These heterogeneous system domains differ significantly in terms of interaction paradigms, communication protocols, and data representation models, provided by supporting middleware platforms. Specifically considering interaction paradigms, the client/server (CS), publish/subscribe (PS), and tuple space (TS) paradigms are among the most widely employed ones today, with numerous related middleware platforms. In light of the above, we have aimed at eliciting abstractions that (i) leverage the diversity of interaction paradigms associated with today's and future complex distributed systems, as well as (ii) enable cross-paradigm interaction to sustain interoperability in the highly heterogeneous Future Internet.

Existing cross-domain interoperability efforts are based on bridging communication protocols, wrapping systems behind standard technology interfaces, and/or providing common API abstractions. In particular, such techniques have been applied by the two widely established system integration paradigms, that is, service oriented architecture (SOA) and enterprise service bus (ESB). However, state of the art interoperability efforts do not or only poorly address interaction paradigm interoperability. Indeed, systems integrated via SOA and ESB solutions have their interaction semantics transformed to the CS paradigm. Then, potential loss of interaction semantics can result in suboptimal or even problematic system integration. To overcome the limitation of today's ESB-based connectors for cross-domain interoperability in the Future Internet, we introduce a new connector type, called GA connector, which stands for "Generic Application connector". The proposed connector type is based on the service bus paradigm in that it achieves bridging across heterogeneous connector types. However, the behavior of the GA connector type differs from that of classical ESB connectors by bridging protocols across heterogeneous paradigms, which is further realized by paying special attention to the preservation of the semantics of the composed protocols. Indeed, the GA connector type is based on the abstraction and semantic-preserving merging of the common high-level semantics of base interaction paradigms.

Eliciting Interaction Paradigm Abstractions: We introduce a systematic abstraction of interaction paradigms with the following features:

- First, we introduce base CS, PS and TS connector types, which formally characterize today's core interaction paradigms. The proposed types comprehensively cover the essential semantics of the considered paradigms, based on a thorough survey of the related literature and representative middleware instances.
- Then, we further abstract these connector types into a single higher-level one, the GA connector type. GA is a comprehensive connector type based on the abstract union of CS, PS, and TS, where precise identification of the commonalities or similarities between the latter has enabled the optimization of the former. Further, GA preserves by construction the semantics of CS, PS, and TS.
- In more detail, connector types are formally specified in terms of: (i) their API (Application Programming Interface), and (ii) their roles, i.e., the semantics of interaction of the connected component(s) with the environment via the connector. Regarding the latter, the behavioral specification of roles from a middleware perspective relates to specifying the production and consumption of information in the network, while the semantics of the information are abstracted and dealt with at the application layer. The behaviors of the connector roles are then specified using Labeled Transition Systems (LTS). We precisely define the mapping of the roles implemented by the base connector types to/from the corresponding roles of the GA connector type.
- For both the above abstraction transformations, we provide counterpart concretizations, which enable transforming GA connector primitives to CS, PS, or TS connector primitives and then to concrete middleware platforms primitives.
- Furthermore, based on the GA abstraction, we introduce mapping transformations between any pair from the set {CS, PS, TS} via GA. The fine knowledge of CS, PS, and TS semantics, as embedded in GA, enables these mappings to be precise: differing semantics are mapped to each other in such a way that loss of semantics is limited to the minimum. These mappings relate to the definition of the glue process implemented by the GA connector, which defines how a pair of producer and consumer roles coordinates in the environment. The GA glue reconciles consumer and producer roles that may differ with respect to time and space coupling as well as scoping. Hence, GA connectors support interactions among highly heterogeneous services of the Future Internet, and especially across domains.

eXtensible Service Bus: We apply the above connector abstractions to introduce an enhanced bus paradigm, the *eXtensible Service Bus (XSB)*. XSB features richer interaction semantics than common ESB implementations to deal effectively with the increased Future Internet heterogeneity. Moreover, from its very conception, XSB incorporates special consideration for the cross-integration of heterogeneous interaction paradigms. When mapping between such paradigms, special attention is paid to the preservation of interaction semantics. XSB has the following features:

- XSB is an abstract bus that prescribes only the high-level semantics of the common bus protocol. The XSB common bus protocol features GA semantics.
- Heterogeneous systems can be plugged into the XSB by employing binding components that adapt between the native middleware of the deployed system and the common bus protocol. This adaptation is based on the systematic abstractions and mappings discussed above
- XSB, being an abstract bus, can have different implementations. This means that it needs to be complemented with a substrate which at least supports: (1) deployment (i.e., plugging) of various systems on the bus, and (2) a common bus protocol implementing GA semantics. With respect to the latter, we envision that a GA protocol realization may either be designed and built from scratch (still supposing at least an IP-based transport substrate) or be implemented by conveying GA semantics on top of an existing higher-level protocol used as transport carrier. The latter solution can be attractive, as it facilitates GA protocol realizations in different contexts and domains.

We have carried out two realizations of XSB for the CHOReOS project [30], the first on PEtALS ESB and the second on EasyESB. The genericity and modularity of our solution allowed for easily porting from the first implementation to the second one. We support interoperable peer-to-peer interaction among the CS, PS, and TS paradigms and provide templates for systematic and highly facilitated building of binding components for middleware platforms that follow any one of the three paradigms.

6.4. Service Oriented Middleware facing the Challenges of the Internet of Things

Participants: Benjamin Billet, Nikolaos Georgantas, Sara Hachem, Valérie Issarny, Yesid Jarma Alvis, Cong Kinh Nguyen, George Mathioudakis.

In our vision, The Future Internet can be defined as the union and cooperation of the Internet of Content, Internet of Services, Internet of Things, and 3D interactive Internet, supported by an expanding network infrastructure foundation [6]. In ARLES, we chose to pay special attention to the Internet of Things (IoT). IoT is characterized by the integration of large numbers of real-world objects (or “things") onto the Internet, with the aim of turning high-level interactions with the physical world into a matter as simple as is interacting with the virtual world today. As such, two devices that will play a key role in the IoT are *sensors* and *actuators*. In fact, such devices are already seeing widespread adoption in the highly localized systems within our cars, mobile phones, laptops, home appliances, etc. In their current incarnation, however, sensors and actuators are used for little more than low-level inferences and basic services. This is partly due to their highly specialized domains (signal processing, estimation theory, robotics, etc.), which demand application programmers to also be domain experts, and partly due to a glaring lack of interconnectivity between all the different devices. Our work within this domain has been focused on two related directions:

- **Architecture of a Service Oriented Middleware for the Mobile Internet of Things**: Adopting the service-oriented architecture (SOA) approach towards middleware (see § 3.3), is an adequate solution towards addressing the heterogeneity and the unknown network topology issues in the IoT. SOA is commonly used in IoT solutions to abstract *things* or their measurements as services. The service-oriented paradigm decouples the functionalities of things from their hardware information or other technical details, and supports three core functionalities: *Discovery* and *Composition* of, and *Access* to services. Typically, in traditional uses of SOA, even if millions of services are registered, there is no need to select and access them all simultaneously. However, in the IoT, discovery, composition and access are undoubtedly more complicated. In fact, it is unlikely for a single or even a few services to be sufficient when providing real world measurements. In most cases, to accurately represent a real-world feature, a large number of services are selected to provide their measurements, and subsequently, all acquired values should be properly aggregated. As a consequence, discovery will return a large set of accessible services, redundant as they may be. Consumers are then expected to access the numerous providers to acquire their measurements, over which they should know the exact aggregation/fusion logic to apply. Furthermore, such logic requires precise knowledge and understanding of the real world and its governing physics and mathematics laws. Clearly, performing discovery, composition and access tasks as presented above incurs high communication and computation costs and is thus not realistic within the large scale IoT. In light of the above issues, we have been *revisiting the SOA and its interaction patterns* to support better scalability and exempt consumers from directly interacting with providers. Specifically, we introduce a **thing-based SOA** to wrap access and computation activities in a middleware that, unlike traditional SOA middleware, is aware of the real world, its physics and its mathematics rules; this has further led to our initial work on the components of such a middleware.

- **Probabilistic Registration for Large Scale Mobile Participatory Sensing**: An increasingly important component of the Internet of Things are modern smart phones, whose constituent sensors and wireless connectivity make them ideal candidates for *mobile participatory sensing*, which aids in providing increased knowledge about the real world while relying on a large number of mobile devices. Those devices can host different types of sensors incorporated in every aspect of our lives. However, given the increasing number of capable mobile devices, any participatory sensing approach should be, first and foremost, *scalable*. To address this challenge, we present an approach to decrease the participation of (sensing) devices in a manner that does not compromise the accuracy of the real-world information while increasing the efficiency of the overall system. To reduce the number of the devices involved, we present a probabilistic registration approach [20], based on a realistic human mobility model, that allows devices to decide whether or not to register their sensing services depending on the probability of other, equivalent devices being present at the locations of their expected path. We used our techniques as the basis of the design and implementation of a registration middleware, using which mobile devices can base their registration decision. Through experiments performed on real and simulated datasets, we show that our approach scales, while not sacrificing significant amounts of sensing coverage.

Our IoT middleware is currently being used by the industrial partners in the FP7 IP CHOReOS project. Complementary to our research on this service oriented middleware for the Internet of Things, we have also been working on suitable abstractions for enabling easy application development for the IoT, discussed next.

6.5. Composing Applications in the Internet of Things

Participants: Peter Sawyer, Pankesh Patel, Animesh Pathak.

As introduced above, the Internet of Things integrates the physical world with the existing Internet, and is rapidly gaining popularity, thanks to the increased adoption of smart phones and sensing devices. Several IoT applications have been reported in recent research, and we expect to see increased adoption of IoT concepts in the fields of personal health, inventory management, and domestic energy usage monitoring, among others.

An important challenge to be addressed in the domain of IoT is to enable domain experts (health-care professionals, architects, city planners, etc.) to develop applications in their fields rapidly, with minimal support from skilled computer science professionals. Similar challenges have already been addressed in the closely related fields of Wireless Sensor and Actuator Networks (WSANs) and Pervasive/Ubiquitous computing. While the main challenge in the former is the *large scale* of the systems (hundreds to thousands of largely similar nodes, sensing and acting on the environment), the primary concern in the latter has been the *heterogeneity* of nodes and the major role that the user's own interaction with these nodes plays in these systems (cf. the classic "smart home" scenario where the user interacts with a smart display which works together with his refrigerator and toaster). The upcoming field of IoT includes both WSANs as well as smart appliances, in addition to the elements of the "traditional" Internet such as Web and database servers, exposing their functionalities as Web services, etc. Consequently, an ideal application development abstraction of the IoT will allow (domain expert) developers to intuitively specify the rich interactions between the extremely large number of disparate devices in the future Internet of Things [19].

The larger goal of our research is to propose a suitable application development framework which addresses the challenges introduced above. To that end, our work this year covered the following related areas:

- **Multi-stage Model-driven approach for IoT Application Development**: We have proposed a multi-stage model-driven approach for IoT application development based on a precise definition of the role to be played by each stakeholder involved in the process – domain expert, application designer, application developer, device developer, and network manager. The metamodels/abstractions available to each stakeholder are further customized using the inputs provided in the earlier stages by other stakeholders. We have also implemented code-generation and task-mapping techniques to support our approach. Our initial evaluation based on two realistic scenarios shows that the use of our techniques/framework succeeds in improving productivity in the IoT application development process.
- **Revisiting Requirements Engineering (RE) Practices for IoT**: Requirements engineering (RE) has evolved to discover, model, specify and manage the required and desired properties of software systems. Conventional RE makes an assumption that the knowledge from which the requirements will be formulated exists a-priori, even though the knowledge may be fragmentary, distributed and tacit. Thus, although their discovery may take significant effort, the requirements are discoverable using the appropriate RE practices.

  However, the last decade or so has seen the emergence of new types of systems where this assumption does not hold, including the IoT. Conventional RE is ill-equipped to discover, model, specify and manage these systems' requirements because incomplete knowledge of the context under which they must operate is available at design time. While some progress has been made, by (e.g.) maintaining requirements models that support reasoning over context at runtime, the IoT has now emerged to compound the challenge for RE. Drawing on experiences from ubiquitous computing and WSAN domains, in [22] we provided initial insights into how the field of RE needs to evolve in order to address the challenges brought forth by IoT.

We have incorporated our continued research in the above areas into *Srijan* (§ 5.5), which provides an easy-to-use graphical front-end to the various steps involved in developing an application using the ATaG macroprogramming framework.

6.6. Requirements-aware Systems for Self-adaptation under Uncertainty

Participants: Romina Torres, Nelly Bencomo, Valérie Issarny, Peter Sawyer.

The development of software-intensive systems is driven by their requirements. Traditional requirements engineering (RE) methods focus on resolving ambiguities in requirements and advocate specifying requirements in sufficient detail so that the implementation can be checked against them for conformance. In an ideal world, this way of thinking can be very effective. Requirements can be specified clearly, updated as necessary, and evolutions of the software design can be made with the requirements in mind.

Increasingly, however, it is not sufficient to fix requirements statically because they will change at runtime as the operating environment changes. Furthermore, as software systems become more pervasive, there is growing uncertainty about the environment and so requirements changes cannot be predicted at design-time. It is considerations such as these that have led to the development of self-adaptive systems (SASs), which have the ability to dynamically and autonomously reconfigure their behavior to respond to changing external conditions.

The key argument of our research is that current software engineering (SE) methods do not support well the kind of dynamic appraisal of requirements needed by a SAS. definition and structure of requirements is lost as requirements are refined into an implementation. Even in cases where requirements monitoring is explicitly included, high-level system requirements must be manually refined into low-level runtime artefacts during the design process so that they can be monitored. There is a lack of approaches supporting for runtime representation, evolution and assessment of requirements. Currently, the approaches mainly assume that it is possible to predefine and envisage the requirements for the total set of target behaviours. Such estimations and beliefs may not be appropriate, if the system is to recover during execution from unforeseen situations, or adapt dynamically to new environmental conditions or to satisfy new requirements that were not foreseen during development. A self-adaptive system is able, at run time, to satisfy new requirements and behaviors. Our research focuses on approaches to support the runtime representation of requirements that will underpin the way a system can reason and assess them during execution.

Our research has been carried out within the research project Marie Curie Fellowship called Requirements-aware Systems (nickname: Requirements@run.time). The research is based on a new paradigm for SE, called requirements-awareness (also known as requirements reflection), in which requirements are reified as runtime entities. Requirements-awareness allows systems to dynamically reason about themselves at the level of the requirements - in much the same way that architectural reflection currently allows runtime reasoning at the level of software. We believe that requirements-awareness (i.e. requirements reflection) will support the development and management of SASs because it will raise the level of discourse at which a software system is able to reflect upon itself.

In the above context, we have been working on the design and implementation of systems with the ability to dynamically observe and reason about their requirements. The results will contribute towards the development of conceptual foundations, engineering techniques, and computing infrastructure for the access and manipulations of runtime abstractions of requirements. Currently, a prototype for the use of runtime goals has been developed. The RELAX language has been proposed to make requirements more tolerant to environmental uncertainty. Design assumptions, called Claims), are applied as markers of uncertainty that document how design assumptions affect goals. Monitoring Claims at runtime has been used to drive self-adaptation. By monitoring Claims during the execution of the systems, their veracity can be tested. If a Claim is falsified, the effect can be propagated to the system's goal model and an alternative (more suitable) means of goal realization will be selected, resulting in dynamic adaptation of the system to a configuration that better satisfies the goals under the prevailing environmental context.

ASAP Project-Team

6. New Results

6.1. Models and abstractions for distributed systems

This section summarizes the major results obtained by the ASAP team that relate to the foundations of distributed systems.

6.1.1. Efficient shared memory consensus

Participants: Michel Raynal, Julien Stainer.

This work is on an efficient algorithm that builds a consensus object. This algorithm is based on an Ω failure detector (to obtain consensus liveness) and a store-collect object (to maintain its safety). A store-collect object provides the processes with two operations, a store operation which allows the invoking process to deposit a new value while discarding the previous value it has deposited and a collect operation that returns to the invoking process a set of pairs (i, val) where val is the last value deposited by the process p_i. A store-collect object has no sequential specification.

While store-collect objects have been used as base objects to design wait-free constructions of more sophisticated objects (such as snapshot or renaming objects), as far as we know, they have not been explicitly used to built consensus objects. The proposed store-collect-based algorithm, which is round-based, has several noteworthy features. First it uses a single store-collect object (and not an object per round). Second, during a round, a process invokes at most once the store operation and the value val it deposits is a simple pair $\langle r, v\rangle$ where r is a round number and v a proposed value. Third, a process is directed to skip rounds according to its view of the current global state (thereby saving useless computation rounds). Finally, the proposed algorithm benefits from the adaptive wait-free implementations that have been proposed for store-collect objects, namely, the number of shared memory accesses involved in a collect operation is $O(k)$ where k is the number of processes that have invoked the store operation. This makes this new algorithm particularly efficient and interesting for multiprocess programs made up of asynchronous crash-prone processes that run on top of multicore architectures.

6.1.2. A Contention-Friendly, Non-blocking Skip List

Participants: Tyler Crain, Michel Raynal.

This work [27] presents a new non-blocking skip list algorithm. The algorithm alleviates contention by localizing synchronization at the least contended part of the structure without altering consistency of the implemented abstraction. The key idea lies in decoupling a modification to the structure into two stages: an eager abstract modification that returns quickly and whose update affects only the bottom of the structure, and a lazy selective adaptation updating potentially the entire structure but executed continuously in the background. As non-blocking skip lists are becoming appealing alternatives to latch-based trees in modern main-memory databases, we integrated it into a main-memory database benchmark, SPECjbb. On SPECjbb as well as on micro-benchmarks, we compared the performance of our new non-blocking skip list against the performance of the JDK non-blocking skip list. Results indicate that our implementation is up to 2:5 faster than the JDK skip list.

6.1.3. STM Systems: Enforcing Strong Isolation between Transactions and Non-transactional Code

Participants: Tyler Crain, Eleni Kanellou, Michel Raynal.

Transactional memory (TM) systems implement the concept of an atomic execution unit called a transaction in order to discharge programmers from explicit synchronization management. But when shared data is atomically accessed by both transaction and non-transactional code, a TM system must provide strong isolation in order to overcome consistency problems. Strong isolation enforces ordering between non-transactional operations and transactions and preserves the atomicity of a transaction even with respect to non-transactional code. This work [29] presents a TM algorithm that implements strong isolation with the following features: (a) concurrency control of non-transactional operations is not based on locks and is particularly efficient, and (b) any non-transactional read or write operation always terminates (there is no notion of commit/abort associated with them).

6.1.4. A speculation-friendly binary search tree

Participants: Tyler Crain, Michel Raynal.

In this work [26], in collaboration with Vincent Gramoli, we introduce the first binary search tree algorithm designed for speculative executions. Prior to this work, tree structures were mainly designed for their pessimistic (non-speculative) accesses to have a bounded complexity. Researchers tried to evaluate transactional memory using such tree structures whose prominent example is the red-black tree library developed by Oracle Labs that is part of multiple benchmark distributions. Although well-engineered, such structures remain badly suited for speculative accesses, whose step complexity might raise dramatically with contention. We show that our speculation-friendly tree outperforms the existing transaction-based version of the AVL and the red-black trees. Its key novelty stems from the decoupling of update operations: they are split into one transaction that modifies the abstraction state and multiple ones that restructure its tree implementation in the background. In particular, the speculation-friendly tree is shown correct, reusable and it speeds up a transaction-based travel reservation application by up to 3:5.

6.1.5. Towards a universal construction for transaction-based multiprocess programs

Participants: Tyler Crain, Damien Imbs, Michel Raynal.

The aim of a Software Transactional Memory (STM) system is to discharge the programmer from the explicit management of synchronization issues. The programmer's job resides in the design of multiprocess programs in which processes are made up of transactions, each transaction being an atomic execution unit that accesses concurrent objects. The important point is that the programmer has to focus her/his efforts only on the parts of code which have to be atomic execution units without worrying on the way the corresponding synchronization has to be realized. Non-trivial STM systems allow transactions to execute concurrently and rely on the notion of commit/abort of a transaction in order to solve their conflicts on the objects they access simultaneously. In some cases, the management of aborted transactions is left to the programmer. In other cases, the underlying system scheduler is appropriately modified or an underlying contention manager is used in order that each transaction be ("practically always" or with high probability) eventually committed. This work [28] presents a deterministic STM system in which (1) every invocation of a transaction is executed exactly once and (2) the notion of commit/abort of a transaction remains unknown to the programmer. This system, which imposes restriction neither on the design of processes nor or their concurrency pattern, can be seen as a step in the design of a deterministic universal construction to execute transaction-based multiprocess programs on top of a multiprocessor. Interestingly, the proposed construction is lock-free (in the sense that it uses no lock).

6.1.6. A Tight RMR Lower Bound for Randomized Mutual Exclusion

Participant: George Giakkoupis.

The Cache Coherent (CC) and the Distributed Shared Memory (DSM) models are standard shared memory models, and the Remote Memory Reference (RMR) complexity is considered to accurately predict the actual performance of mutual exclusion algorithms in shared memory systems. Through a collaboration with Philipp Woelfel [32], we proved a tight lower bound for the RMR complexity of deadlock-free randomized mutual exclusion algorithms in both the CC and the DSM model with atomic registers and compare&swap objects and an adaptive adversary. Our lower bound establishes that an adaptive adversary can schedule n processes in such a way that each enters the critical section once, and the total number of RMRs is $\Omega(n \log n / \log \log n)$ in expectation. This matches an upper bound of Hendler and Woelfel (2011).

6.1.7. On the Time and Space Complexity of Randomized Test-And-Set

Participant: George Giakkoupis.

Through a collaboration with Philipp Woelfel [33] we studied the time and space complexity of randomized Test-And-Set (TAS) implementations from atomic read/write registers in asynchronous shared memory models with n processes. We presented an adaptive TAS implementation with an expected (individual) step complexity of $O(\log^* k)$, for contention k, against the oblivious adversary, improving a previous (non-adaptive) upper bound of $O(\log \log n)$ by Alistarh and Aspnes (2011). We also showed how to modify the adaptive RatRace TAS algorithm by Alistarh, Attiya, Gilbert, Giurgiu, and Guerraoui (2010) to improve the space complexity from $O(n^3)$ to $O(n)$, while maintaining logarithmic expected step complexity against the adaptive adversary. Finally, we proved that for any randomized 2-process TAS algorithm there exists a schedule determined by an oblivious adversary, such that with probability at least $1/4^t$ one of the processes does not finish its TAS operation in within fewer than t steps. This complements a lower bound by Attiya and Censor-Hillel (2010) of a similar result for $n \geq 3$ processes.

6.2. Large-scale and user-centric distributed system

6.2.1. WhatsUp: P2P news recommender

Participants: Antoine Boutet, Davide Frey, Arnaud Jegou, Anne-Marie Kermarrec.

The main application in the context of GOSSPLE is WhatsUp, an instant news system designed for a large-scale network with no central authority. WhatsUp builds an implicit social network based on the opinions users express about the news items they receive (like-dislike). This is achieved through an obfuscation mechanism that does not require users to ever reveal their exact profiles. WhatsUp disseminates news items through a novel heterogeneous gossip protocol that biases the choice of its targets towards those with similar interests and amplifies dissemination based on the level of interest in every news item. WhatsUp outperforms various alternatives in terms of accurate and complete delivery of relevant news items while preserving the fundamental advantages of standard gossip: namely simplicity of deployment and robustness. This work has been carried out in collaboration with Rachid Guerraoui from EPFL and was demonstrated during the different local events and will appear in IPDPS 2013 [21].

6.2.2. Privacy in P2P recommenders

Participants: Antoine Boutet, Davide Frey, Arnaud Jegou, Anne-Marie Kermarrec.

We also propose a mechanism to preserve privacy in WhatsUp, which can also be used in any distributed recommendation system. Our approach relies on (i) an original obfuscation mechanism hiding the exact profiles of users without significantly decreasing their utility, as well as (ii) a randomized dissemination algorithm ensuring differential privacy during the dissemination process. Results show that our solution preserves accuracy without the need for users to reveal their preferences. Our approach is also flexible and robust to censorship.

6.2.3. BLIP: Non-interactive differentially-private similarity computation on Bloom filters

Participants: Mohammad Alaggan, Anne-Marie Kermarrec.

In this project [19], done in collaboration with Sébastien Gambs (team CIDRE), we consider the scenario in which the profile of a user is represented in a compact way, as a Bloom filter, and the main objective is to privately compute in a distributed manner the similarity between users by relying only on the Bloom filter representation. In particular, we aim at providing a high level of privacy with respect to the profile even if a potentially unbounded number of similarity computations take place, thus calling for a non-interactive mechanism. To achieve this, we propose a novel non-interactive differentially private mechanism called BLIP (for BLoom-and-flIP) for randomizing Bloom filters. This approach relies on a bit flipping mechanism and offers high privacy guarantees while maintaining a small communication cost. Another advantage of this non-interactive mechanism is that similarity computation can take place even when the user is offline, which is impossible to achieve with interactive mechanisms. Another of our contributions is the definition of a probabilistic inference attack, called the “Profile Reconstruction Attack”, that can be used to reconstruct the profile of an individual from his Bloom filter representation, along with the “Profile Distinguishing Game”. More specifically, we provide an analysis of the protection offered by BLIP against this profile reconstruction attack by deriving an upper and lower bound for the required value of the differential privacy parameter ϵ.

6.2.4. Heterogeneous Differential Privacy

Participants: Mohammad Alaggan, Anne-Marie Kermarrec.

The massive collection of personal data by personalization systems has rendered the preservation of privacy of individuals more and more difficult. Most of the proposed approaches to preserve privacy in personalization systems usually address this issue uniformly across users, thus completely ignoring the fact that users have different privacy attitudes and expectations (even among their own personal data). In this project, in collaboration with Sébastien Gambs (team CIDRE), we propose to account for this non-uniformity of privacy expectations by introducing the concept of heterogeneous differential privacy. This notion captures both the variation of privacy expectations among users as well as across different pieces of information related to the same user. We also describe an explicit mechanism achieving heterogeneous differential privacy, which is a modification of the Laplacian mechanism due to Dwork [54], we evaluate on real datasets the impact of the proposed mechanism with respect to a semantic clustering task. The results of our experiments clearly demonstrate that heterogeneous differential privacy can account for different privacy attitudes while sustaining a good level of utility as measured by the recall.

6.2.5. Social Market

Participants: Davide Frey, Arnaud Jegou, Anne-Marie Kermarrec, Michel Raynal, Julien Stainer.

The ability to identify people that share one’s own interests is one of the most interesting promises of the Web 2.0 driving user-centric applications such as recommendation systems or collaborative marketplaces. To be truly useful, however, information about other users also needs to be associated with some notion of trust. Consider a user wishing to sell a concert ticket. Not only must she find someone who is interested in the concert, but she must also make sure she can trust this person to pay for it. Social Market (SM) solve this problem by allowing users to identify and build connections to other users that can provide interesting goods or information and that are also reachable through a trusted path on an explicit social network like Facebook. This year, we extended the contributions presented in 2011, by introducing two novel distributed protocols that combine interest-based connections between users with explicit links obtained from social networks a-la Facebook. Both protocols build trusted multi-hop paths between users in an explicit social network supporting the creation of semantic overlays backed up by social trust. The first protocol, TAPS2, extends our previous work on *TAPS* (Trust-Aware Peer Sampling), by improving the ability to locate trusted nodes. Yet, it remains vulnerable to attackers wishing to learn about trust values between arbitrary pairs of users. The second protocol, PTAPS (*Private TAPS*), improves TAPS2 with provable privacy guarantees by preventing users from revealing their friendship links to users that are more than two hops away in the social network. In addition to proving this privacy property, we evaluate the performance of our protocols through event-based simulations, showing significant improvements over the state of the art. We submitted this work for journal publication.

6.2.6. Geolocated Social Networks

Participants: Anne-Marie Kermarrec, François Taïani.

Geolocated social networks, that combine traditional social networking features with geolocation information, have grown tremendously over the last few years. Yet, very few works have looked at implementing geolocated social networks in a fully distributed manner, a promising avenue to handle the growing scalability challenges of these systems. In [25], we have focused on georecommendation, and showed that existing decentralized recommendation mechanisms perform in fact poorly on geodata. In this work, we have proposed a set of novel gossip-based mechanisms to address this problem, and captured these mechanisms in a modular similarity framework called "Geology". The resulting platform is lightweight, efficient, and scalable. More precisely, we have shown its benefits in terms of recommendation quality and communication overhead on a real data set of 15,694 users from Foursquare, a leading geolocated social network.

6.2.7. Content and Geographical Locality in User-Generated Content Sharing Systems

Participants: Anne-Marie Kermarrec, Konstantinos Kloudas, François Taïani.

User Generated Content (UGC), such as YouTube videos, accounts for a substantial fraction of the Internet traffic. To optimize their performance, UGC services usually rely on both proactive and reactive approaches that exploit spatial and temporal locality in access patterns. Alternative types of locality are also relevant and hardly ever considered together. In [34], we show on a large (more than $650,000$ videos) YouTube dataset that content locality (induced by the related videos feature) and geographic locality, are in fact correlated. More specifically, we show how the geographic view distribution of a video can be inferred to a large extent from that of its related videos. We leverage these findings to propose a UGC storage system that proactively places videos close to the expected requests. Compared to a caching-based solution, our system decreases by 16% the number of requests served from a different country than that of the requesting user, and even in this case, the distance between the user and the server is 29% shorter on average.

6.2.8. Probabilistic Deduplication for Cluster-Based Storage Systems

Participants: Davide Frey, Anne-Marie Kermarrec, Konstantinos Kloudas.

The need to backup huge quantities of data has led to the development of a number of distributed deduplication techniques that aim to reproduce the operation of centralized, single-node backup systems in a cluster-based environment. At one extreme, stateful solutions rely on indexing mechanisms to maximize deduplication. However the cost of these strategies in terms of computation and memory resources makes them unsuitable for large-scale storage systems. At the other extreme, stateless strategies store data blocks based only on their content, without taking into account previous placement decisions, thus reducing the cost but also the effectiveness of deduplication. In [30], we propose, Produck, a stateful, yet lightweight cluster-based backup system that provides deduplication rates close to those of a single-node system at a very low computational cost and with minimal memory overhead. In doing so, we provide two main contributions: a lightweight probabilistic node-assignment mechanism and a new bucketbased load-balancing strategy. The former allows Produck to quickly identify the servers that can provide the highest deduplication rates for a given data block. The latter efficiently spreads the load equally among the nodes. Our experiments compare Produck against state-of-the-art alternatives over a publicly available dataset consisting of 16 full *Wikipedia* backups, as well as over a private one consisting of images of the environments available for deployment on the Grid5000 experimental platform. Our results show that, on average, Produck provides (i) up to 18% better deduplication compared to a stateless minhash-based technique, and (ii) an 18-fold reduction in computational cost with respect to a stateful BloomFilter-based solution.

6.2.9. Large scale analysis of HTTP adaptive streaming in mobile networks

Participants: Ali Gouta, Anne-Marie Kermarrec.

In collaboration with Yannick Le Louedec and Nathalie Amann we have been working in the context of adaptive streaming in mobile networks. HTTP Adaptive bitrate video Streaming (HAS) is now widely adopted by Content Delivery Network Providers (CDNPs) and Telecom Operators (Telcos) to improve user Quality of Experience (QoE). In HAS, several versions of videos are made available in the network so that the quality of the video can be chosen to better fit the bandwidth capacity of users. These delivery requirements raise new challenges with respect to content caching strategies, since several versions of the content may compete to be cached. We used a real HAS dataset collected in France and provided by a mobile telecom operator involving more than 485,000 users requesting adaptive video contents through more than 8 million video sessions over a 6 week measurement period. Firstly, we proposed a fine-grained definition of content popularity by exploiting the segmented nature of video streams. We also provided analysis about the behavior of clients when requesting such HAS streams. We proposed novel caching policies tailored for chunk-based streaming. Then we studied the relationship between the requested video bitrates and radio constraints. Finally, we studied the users' patterns when selecting different bitrates of the same video content. Our findings provide useful insights that can be leveraged by the main actors of video content distribution to improve their content caching strategy for adaptive streaming contents as well as to model users' behavior in this context.

6.2.10. Regenerating Codes: A System Perspective

Participants: Anne-Marie Kermarrec, Alexandre van Kempen.

The explosion of the amount of data stored in cloud systems calls for more efficient paradigms for redundancy. While replication is widely used to ensure data availability, erasure correcting codes provide a much better trade-off between storage and availability. Regenerating codes are good candidates for they also offer low repair costs in term of network bandwidth. While they have been proven optimal, they are difficult to understand and parameterize. In collaboration with Nicolas Le Scouarnec, Gilles Straub and Steve Jiekak from Technicolor, we performed an analysis of regenerating codes, which enables practitioners to grasp the various trade-offs. More specifically we made two contributions: (i) we studied the impact of the parameters by conducting an analysis at the level of the system, rather than at the level of a single device; (ii) we compared the computational costs of various implementations of codes and highlight the most efficient ones. Our goal is to provide system designers with concrete information to help them choose the best parameters and design for regenerating codes.

6.2.11. Availability-based methods for distributed storage systems

Participants: Anne-Marie Kermarrec, Alexandre van Kempen.

Distributed storage systems rely heavily on redundancy to ensure data availability as well as durability. In networked systems subject to intermittent node unavailability, the level of redundancy introduced in the system should be minimized and maintained upon failures. Repairs are well- known to be extremely bandwidth-consuming and it has been shown that, without care, they may significantly congest the system. In collaboration with Gilles Straub and Erwan Le Merrer from Technicolor, we proposed an approach to redundancy management accounting for nodes heterogeneity with respect to availability. We show that by using the availability history of nodes, the performance of two important faces of distributed storage (replica placement and repair) can be significantly improved. Replica placement is achieved based on complementary nodes with respect to nodes availability, improving the overall data availability. Repairs can be scheduled thanks to an adaptive per-node timeout according to node availability, so as to decrease the number of repairs while reaching comparable availability. We propose practical heuristics for those two issues. We evaluate our approach through extensive simulations based on real and well-known availability traces. Results clearly show the benefits of our approach with regards to the critical trade-off between data availability, load-balancing and bandwidth consumption.

6.2.12. On The Impact of Users Availability In OSNs

Participants: Antoine Boutet, Anne-Marie Kermarrec, Alexandre van Kempen.

Availability of computing resources has been extensively studied in literature with respect to uptime, session lengths and inter-arrival times of hardware devices or software applications. Interestingly enough, information related to the presence of users in online applications has attracted less attention. Consequently, only a few attempts have been made to leverage user availability pattern to improve such applications. In collaboration with Erwan Le Merrer from Technicolor, we studied an availability trace collected from MySpace. Our results show that the online presence of users tends to be correlated to that of their friends. User availability also plays an important role in some algorithms and focus on information spreading. In fact, identifying central users i.e. those located in central positions in a network, is key to achieve a fast dissemination and the importance of users in a social graph precisely vary depending on their availability.

6.2.13. Chemical programming model

Participant: Marin Bertier.

This work, done in collaboration with the Myriads project team, focuses on chemical programming, a promising paradigm to design autonomic systems. The metaphor envisions a computation as a set of concurrent reactions between molecules of data arising non-deterministically, until no more reactions can take place, in which case, the solution contains the final outcome of the computation.

More formally, such models strongly rely on concurrent multiset rewriting: the data are a multiset of molecules, and reactions are the application of a set of conditioned rewrite rules. At run time, these rewritings are applied concurrently, until no rule can be applied anymore (the elements they need do not exist anymore in the multiset). One of the main barriers towards the actual adoption of such models come from their complexity at run time: each computation step may require a complexity in $O(n^k)$ where n denotes the number of elements in the multiset, and k the size of the subset of elements needed to trigger one rule.

Our objective is to design a distributed chemical platform implementing such concepts. This platform should be adapted to large scale distributed system to benefit at his best the inherent distribution of chemical program.

Within this context, we proposed a protocol for the atomic capture of objects in a DHT [20]. This protocol is distributed and evolving over a large scale platform. As the density of potential request has a significant impact on the liveness and efficiency of such a capture, the protocol proposed is made up of two sub-protocols, each of them aimed at addressing different levels of densities of potential reactions in the solution. While the decision to choose one or the other is local to each node participating in a program's execution, a global coherent behavior is obtained.

ASCOLA Project-Team

6. New Results

6.1. Software composition

Participants: Akram Ajouli, Diana Allam, Omar Chebaro, Rémi Douence, Hervé Grall, Jean-Claude Royer, Mario Südholt.

We have produced results on service-oriented computing, language support for software composition, program transformation for composition, as well as the analysis of C programs.

6.1.1. Program transformation and formal properties

We have proposed an extension of the type theory underlying the Coq theorem prover and studied invertible transformations as a means to decompose object-oriented properties.

6.1.1.1. Forcing in the Calculus of Constructions and Coq

We have developed an intuitionistic forcing translation for the Calculus of Constructions (CoC), a translation that corresponds to an internalization of the presheaf construction in CoC [22]. Depending on the chosen set of forcing conditions, the resulting type theory can be extended with extra logical principles. The translation is proven correct—in the sense that it preserves type checking—and has been implemented in Coq. As a case study, we have shown how the forcing translation on integers (which corresponds to the internalization of the topos of trees) allows us to define general inductive types in Coq, without the strict positivity condition.

6.1.1.2. Invertible transformations for program decompositions

When one chooses a main axis of structural decomposition for a software, such as function- or data-oriented decompositions, the other axes become secondary, which can be harmful when one of these secondary axes becomes of main importance. In the context of modular maintenance, we have tackled this problem using invertible program transformations [19]. We have experimented our approach for Java [29] and Haskell programs.

In [29] we have presented such a transformation for Java. Precisely, we build a reversible transformation between Composite and Visitor design patterns in Java programs, based on chains of basic refactoring operations. Such transformations represent an automatic reversible switching between different program architectures with a guarantee of semantic preservation. The transformation is automated with the refactoring tool of a popular IDE: JetBrains Intellij Idea.

As seen in that paper, basic refactoring operations can be combined to perform complex program transformations. But the resulting composed operations are rarely reused, even partially, because popular tools have few support for composition. In [45] we have formalized the composition of refactoring operations of our Composite/Visitor transformation by the means of a static type system. That type system is based on two previous calculi for composition of refactoring, which we recast in one single calculus. The type system is used to prove non-failure and correctness of transformations. This kind of formalization yields a validation of transformations and, if integrated in existing IDEs, should help to reuse existing transformations.

6.1.2. Service-oriented computing

In the field of service-oriented computing, we have developed three contributions: a model for web services that enables WS*/SOAP-based heavyweight services and RESTful lightweight services to be handled uniformly, a type system that is safe in the presence of malicious agents and insecure communication channels, as well as a pivot language that provides a common abstraction for very different web query languages.

6.1.2.1. Uniform service model

Service-oriented applications are frequently used in highly dynamic contexts: service compositions may change dynamically, in particular, because new services are discovered at runtime. Moreover, subtyping has recently been identified as a strong requirement for service discovery. Correctness guarantees over service compositions, provided in particular by type systems, are highly desirable in this context. However, while service oriented applications can be built using various technologies and protocols, none of them provides decent support ensuring that well-typed services cannot go wrong. Currently, Service-Oriented Architecture applications are typically built using either the SOAP/WS or REST service models. Although there is a clear need for a model integrating both in multiple real-world contexts, no integrated model does (yet) exist. Therefore, in [15] we have introduced a model as a foundation for heterogeneous services, therefore unifying the SOAP/WSand RESTmodels.

6.1.2.2. A type system for services

We have presented a formal model in [14] for service compositions and defined a type system [33] with subtyping that ensures type soundness by combining static and dynamic checks. Our model allows channel mobility and inference of the type of discovered channels. This type system is based on the notion of semantic typing and proved to be sound. We have also demonstrated how to get type soundness in presence of malicious agents and insecure communication channels.

6.1.2.3. Criojo: a pivot language for services

Interoperability remains a significant challenge in service-oriented computing. After proposing a pivot architecture to solve three interoperability problems, namely adaptation, integration and coordination problems between clients and servers, we explore the theoretical foundations for this architecture. A pivot architecture requires a universal language for orchestrating services and a universal language for interfacing resources. Since there is no evidence today that Web Services technologies can provide this basis, we have proposed a new language called Criojo and shown that it can be considered as a pivot language. We have formalized the language Criojo and its operational semantics by resorting to a chemical abstract machine, and given an account of formal translations into Criojo: in a distributed context, we have dealt with idiomatic languages for four major programming paradigms: imperative programming, logic programming, functional programming and concurrent programming.

6.1.3. Languages and composition models

We have contributed new results in the domains of software product lines, model-based composition and language support for numerical constraint-based programming.

6.1.3.1. Software product lines and model composition

Many approaches to creating Software Product Lines have emerged that are based on Model-Driven Engineering. Our book [32] introduces both Software Product Lines and Model-Driven Engineering, which have separate success stories in industry, and focuses on the practical combination of them. It describes the challenges and benefits of merging these two software development trends and provides the reader with a novel approach and practical mechanisms to improve variability. Advanced concepts like fine-grained variability and decision models based on aspect-oriented programming techniques are illustrated. The concepts and methods are detailed with two product line examples: the classic smart-home systems and a collection manager information system.

6.1.3.2. Expressive language support for numerical constraint based programming

A combinatorial search can either be performed by using an implicit or an explicit search tree. We have proposed a functional DSL [35] for explicit search trees in the field of numerical constraints. The first advantage of our approach is expressiveness: we can write new algorithms or reformulate existing ones in a simple and unified way. The second advantage is efficiency, since an implicit search may also lead to a blowup of redundant computations. We illustrate this through various examples.

6.1.4. Analysis and test of C programs

Ascola members have participated, in cooperation with researchers from CEA List institute, in the development of analyses and corresponding tool support for C programs.

We have studied combinations of static and dynamic analysis techniques that enable the detection of out-of-bounds memory accesses in C programs and generate corresponding concrete test data [17]. This is particularly problematic for input arrays and pointers in C functions. We have presented a specific technique allowing the interpretation and execution of assertions involving the size of an input array (pointer) of a C function. We have successfully applied this technique in the Sante tool from the CEA where it allows potential out-of-bounds access errors to be detected and classified in several real-life programs.

PathCrawler is a test generation tool developed at CEA LIST for structural testing of C programs. The new version of PathCrawler [18] we have contributed to is developed in an entirely novel form: that of a test-case generation web service which is freely accessible at PathCrawler-online.com. This service allows many test-case generation sessions to be run in parallel in a completely robust and secure way. We have presented PathCrawler-online.com in the form of a lesson on structural software testing, showing its benefits, limitations and illustrating the usage of the tool on a series of examples.

6.2. Aspect-Oriented Programming

Participants: Rémi Douence, Guilhem Jaber, Ismael Mejía, Jacques Noyé, Mario Südholt, Nicolas Tabareau.

We have contributed to the foundations of aspect-oriented programming and presented new programming languages for aspects and related paradigms.

6.2.1. Formal models for AOP

We have presented two calculi contributing to the foundations of AOP: the A Calculus, a parameterized calculus encompassing AspectJ-like and history based aspect languages, and a category-theoretic definition of AOP in terms of 2-categories.

6.2.1.1. The A Calculus

With partners from Vrije Universiteit Brussel and Aarhus University, we have extended the foundational calculus for AOP (introduced in 2010) that supports the most general aspect model to-date compared to existing calculi and the deepest integration with plain OO concepts [12]. Integration with OOP is achieved essentially by modeling proceed using first-class closures. Two well-known pointcut categories, call and execution that are commonly considered similar are shown to be significantly different; our calculus enables the resolution of the associated soundness problems. The A-calculus also includes type ranges, an intuitive and concise alternative to explicit type variables that allows advices to be polymorphic over intercepted methods. Finally, our calculus is the first aspect calculus to use calculus parameters to cover type safety for a wide design space of other features. The soundness of the resulting type system has been verified in Coq.

In 2012, we have covered a range of choices with respect to evaluation order and non-determinism. We have studied one version that enforces a deterministic call-by-value semantics, and another one that omits restrictions on evaluation order and allows many kinds of non-determinism. Furthermore, we have provided a mechanized complete type soundness proof using the theorem prover Coq.

6.2.1.2. A category-theoretic foundation of aspects

Aspect-Oriented Programming (AOP) started fifteen years ago with the remark that modularization of so-called crosscutting functionalities is a fundamental problem for the engineering of large-scale applications. However, theoretical foundations of AOP have been much less studied than its applicability.

We have proposed [26] to put a bridge between AOP and the notion of 2-category to enhance the conceptual understanding of AOP. Starting from the connection between the λ-calculus and the theory of categories, we have defined an internal language for 2-categories and shown how it can be used to define the first categorical semantics for a realistic functional AOP language. We have then used this categorical framework to introduce the notion of computational 2-monads for AOP. We have illustrated their conceptual power by defining a 2-monad for Éric Tanter's execution levels—which constitutes the first algebraic semantics for execution levels—and then introducing the first exception monad transformer specific to AOP that gives rise to a non-flat semantics for exceptions by taking levels into account.

6.2.2. Programming languages for aspects and related paradigms

We have introduced three results related to aspect-based programming languages: an extension of EScala for multi-paradigm programming; Monascheme, a language for modular prototyping of aspect-based languages and language support for membranes, an aspect-based means for structuring computations.

6.2.2.1. Concurrent multi-paradigm programming with EScala

EScala integrates, around the notion of *declarative events*, object-oriented, aspect-oriented and event-based programming [30]. However, in spite of the fact that events naturally evoke some form of concurrency, there is no specific support for concurrency in EScala. It is up to the programmer to understand how to combine declarative events and Scala's support for concurrent programming. We have started working on injecting concurrency at the heart of declarative events so that events can indeed be naturally concurrent [28].

6.2.2.2. Monascheme: modular prototyping of aspect languages

We have then developed Monascheme [21], an extensible aspect-oriented programming language based on monadic aspect weaving. Extensions to the aspect language are defined as monads, enabling easy, simple and modular prototyping. The language is implemented as an embedded language in Racket. We illustrate the approach with an execution level monad and a level-aware exception transformer. Semantic variations can be obtained through monad combinations.

6.2.2.3. Structuring computations with aspect-based membranes

In most aspect-oriented languages, aspects have an unrestricted global view of computation. Several approaches for aspect scoping and more strongly encapsulated modules have been formulated to restrict the power of aspects. Our approach [27] leverages the concept of programmable membranes of Boudol, Schmitt and Stefani, as a means to tame aspects by customizing the semantics of aspect weaving locally. Membranes have the potential to subsume previous proposals in a uniform framework. Because membranes give structure to computation, they enable flexible scoping of aspects; because they are programmable, they enable visibility and safety constraints, both for the advised program and for the aspects. The power and simplicity of membranes open interesting perspectives to unify multiple approaches that tackle the unrestricted power of aspects.

6.3. Cloud applications and infrastructures

Participants: Frederico Alvares, Gustavo Bervian Brand, Yousri Kouki, Adrien Lèbre, Thomas Ledoux, Guillaume Le Louët, Jean-Marc Menaud, Jonathan Pastor, Rémy Pottier, Flavien Quesnel, Mario Südholt.

We have contributed on SLA management for Cloud elasticity, fully distributed and autonomous virtual machine scheduling, and energy-efficient Cloud infrastructures.

6.3.1. SLA Management for Cloud elasticity

In [23], we have introduced *CSLA*, the Cloud Service Level Agreement language. The CSLA language has been influenced by related work, in particular WSLA and SLA@SOI. It allows to describe the SLA between a cloud service provider and a cloud customer. One of the novelties of CSLA is that it integrates features dealing with QoS uncertainty and cloud fluctuations, such as *confidence*, *penalty* and *fuzziness*.

Cloud computing is a model for enabling on-demand access to a shared pool of configurable resources as services. However, the management of such elastic resources is a complex issue. In [24], we have proposed a SLA-driven approach for optimizing the resources capacity planning for Cloud applications. We have modeled Cloud services using a closed queuing network model and proposed an extension of a Mean Value Analysis (MVA) algorithm to take into account the concept of SLA. Then, based on capacity planning method, our solution calculates the optimal configuration of a Cloud application.

6.3.2. Fully distributed and autonomous virtualized environments

Extending previous preliminary results of the DVMS prototype, we have consolidated this system to obtain a fully distributed virtual machine scheduler [13]. This system makes it possible to schedule VMs cooperatively and dynamically in large scale distributed systems. Simulations (up to 64K VMs) and real experiments both conducted on the Grid'5000 large-scale distributed system [34] showed that DVMS is scalable. This building block is a first element of a more complete cloud OS, entitled DISCOVERY (DIStributed and COoperative mechanisms to manage Virtual EnviRonments autonomicallY) [66]. The ultimate goal of this system is to overcome the main limitations of the traditional server-centric solutions. The system, currently under investigation in the context of the Jonathan Pastor's PhD, relies on a peer-to-peer model where each agent can efficiently deploy, dynamically schedule and periodically checkpoint the virtual environments it manages.

6.3.3. Energy-efficient Cloud applications and infrastructures

As a direct consequence of the increasing popularity of Cloud Computing solutions, data centers are amazingly growing and hence have to urgently face with the energy consumption issue. Available solutions rely on Cloud Computing models and virtualization techniques to scale up/down application based on their performance metrics. Although those proposals can reduce the energy footprint of applications and by transitivity of cloud infrastructures, they do not consider the internal characteristics of applications to finely define a trade-off between applications Quality of Service and energy footprint. We have proposed a self-adaptation approach that considers both application internals and system to reduce the energy footprint in cloud infrastructure [31], [11]. Each application and the infrastructure are equipped with their own control loop, which allows them to autonomously optimize their executions. In addition, these autonomic loops are coordinated to avoid inconsistent states. This coordination improves the synergy between applications and infrastructure in order to optimize the infrastructure energy consumption [16].

We have extended our previous work on Entropy, a virtual machine placement manager, by the development of btrScript, a safe autonomic system for virtual machine management that includes actions and placement rules. Actions are imperative operations to reconfigure the data center and declarative rules specify the virtual machine placement. Administrators schedule both actions and rules, to manage their data center(s). They can also interact with the btrScript system in order to monitor the data center and compute the correct virtual machine placement [25]. btrScript and Entropy have been packaged in a common software btrCloud.

ATLANMOD Team

6. New Results

6.1. Core Modeling technologies

AtlanMod has continued to improve the core model and model transformation technologies that are reused in the other more domain-oriented research lines of the team. Main results in this area have been:

- Model-to-Model Transformation Refactorings. In object-oriented programming, continuous refactorings are used as the main mechanism to increase the maintainability of the code base. Unfortunately, in the field of model transformations, such refactoring support is so far missing. We have tackled this limitation by adapting the notion of refactorings to model-to-model (M2M) transformations. Particularly, in [17] we present a dedicated catalogue of refactorings for improving the quality of M2M transformations. This catalogue is the result of analyzing existing ATL transformations; its scope is beyond ATL, covering other M2M transformation languages.
- Reactive Model Transformations. Model-driven applications manipulate models by executing model transformations that are seen by host applications as black-box functions returning the computed target models. We propose a paradigm shift where a network of reactive transformations defines persistent data-flows among models. A reactive transformation engine takes care of activating only the strictly needed computation in response to updates or requests of model elements. Computation is updated when necessary, in an autonomous and optimized way. The application architecture results deeply changed, since the host application does not directly control the execution of the transformations anymore, but only accesses or updates the underlying models. We experiment this paradigm by implementing a reactive engine for ATL.
- EMF Profiles. There are many situations in which one needs to extend or annotate a model with additional information. Nevertheless, changing the metamodel to include this new information is very costly (e.g. you'll need to recreate the modeling environment and, possibly, to migrate other existing models). As a solution, we have proposed the idea of EMF Profiles [16] as a way to reuse the idea of UML Profiles for general EMF Models. UML profiles have been a key enabler for the success of UML by providing a lightweight language-inherent extension mechanism which is expressive enough to cover an important subset of adaptation scenarios. We believe a similar concept for DSMLs would provide an easier extension mechanism for EMF.

6.2. Domain-Specific languages

In the field of Domain-Specific Languages (DSLs), we have focused on the improvement of the DSLs definition process. During 2012 the new results in this area have been:

- Software development processes are becoming more collaborative, trying to integrate end-users as much as possible. The idea is to advance towards a community-driven process where all actors (both technical and nontechnical) work together to ensure that the system-to-be will satisfy all expectations. This seems specially appropriate in the field of Domain-Specific Languages (DSLs) typically designed to facilitate the development of software for a particular domain. We have designed a collaborative infrastructure for the development of DSLs where end-users have a direct and active participation in the evolution of the language [22], [32]. This infrastructure is based on Collaboro, a DSL to represent change proposals, possible solutions and comments arisen during the development and evolution of a language.

- When developing DSLs, a number of design decisions must be made, such as those related to its concrete syntax, how the language semantics is going to be defined and in which form (interpreted or compiled), or whether there will be an underlying abstract syntax. However, deciding whether the DSL will be internal or external will have an impact on the other aspects of the language. Making an effective choice between these two options therefore requires a careful evaluation of the pros and cons of each alternative. Some important aspects that should be evaluated are the following, which are related to the three elements of a DSL: abstract and concrete syntaxes, and semantics (executability and optimizations), and to quality criteria (extensibility and efficiency) and DSL tooling (tools for developing DSL and tools for using DSL). In [40] we presented the results of this work.

6.3. Model Verification

Guranteeing the correctness of models is a very important element of the MDE infrastructure. We made several contributions to the model verification field in 2012:

- Automated verification of declarative, rule-based model transformations. Having sound transformations is essential, as they are the compilers in MDE. Because transformations are created frequently, e.g., on a per-project basis, it is important that we can check their correctness automatically. We have developed a novel, automatic proof technique based on Satisfiability Modulo Theories (SMT) solving [20] for this, as well as a bounded-search verification approach using relational logic and Alloy [21]. Both Yices and Z3 have been used as SMT solvers for this work.
- Improving EMFtoCSP, the AtlanMod model finder. Model finding is a central, recurring task in MDE. It subsumes both metamodel consistency checking (i.e., metamodel verification) and metamodel instantiation (e.g., test case generation). Even when using a bounded search approach, the underlying research problem is computational hard and calls for flexibile solutions and heuristics. We have generalized and improved the EMFtoCSP model finder (formerly: UMLtoCSP), which is based on Constraint Logic Programming (CLP). It now supports both UML and Ecore (and OCL constraints) and is open for further modeling languages [26]. As the first available MDE model finder, it now supports reasoning over string constraints. Such constraints are common in practical applications of MDE, but none of the existing model finding approaches could handle them. We have developed a flexible string constraint solver (based on multi-head constraint handling rules) that seamlessly integrates into EMTtoCSP [19].

6.4. Model Transformation Testing

White-box testing for model transformations is a technique that involves the extraction of knowledge embedded in the transformation code to generate test models. In [31], we manually extract such knowledge and we represent it in the form of partial models that can drive the generation of highly effective test models. In other works we go a step further and use static analysis to automatically extract testing knowledge from transformation code. We propose two tool-supported methodologies to automatically generate test cases using structural information from a model transformation. In [27] we have developed an approach that optimizes the test coverage while testing rule-based model transformation languages like ATL. The approach is based on analyzing the dependencies among the OCL queries that are used within the transformation code. The methodology in [29] makes use of the metamodel footprinting mechanism, generates partial models representing the testing intent and uses the ALLOY solver to create complete usable models. The experimental results show that a limited amount of white-box information on the model transformation (i.e., our footprints) can provide remarkable improvements on the efficiency of the generated tests.

6.5. Reverse Engineering

Model Driven Reverse Engineering (MDRE), and its applications such as software modernization, is a discipline in which model-driven development (MDD) techniques are used to treat legacy systems. During 2012, Atlanmod has continued working actively on this research area. The main contributions are the following:

Grammar-to-Model Bridging When existing software artifacts are treated in MDRE, they must be first transformed into models to apply MDD techniques such as model transformations. Since most scenarios involve dealing with code in general-purpose programming languages (GPL), the extraction of models from GPL code is an essential task. We designed Grammar-to-Model Transformation Language (Gra2MoL) as a domain-specific language (DSL) tailored to the extraction of models from GPL code. Gra2MoL aims to reduce the effort needed to implement grammarware-MDD bridges, since building dedicated parsers is a complex and time-consuming task. The language also provides a powerful query language which eases the retrieval of scattered information in syntax trees. Moreover, it incorporates extensibility and grammar reuse mechanisms. In [13], Gra2MoL is described in detail and a case study based on the application of the language in the extraction of models from Delphi code is included.

API-to-Model Bridging Software systems usually manage many Application Programming Interfaces (APIs) to access different software assets (e.g., databases, middleware, etc). A MDRE process therefore also normally involves extracting models from legacy artifacts using API. Thus, we devised API2MoL [14], a DSL which allows developers defining technological bridges between the model and the API technologies. API2MoL is, to the best of our knowledge, the first generic proposal to deal with the integration of MDE and APIs which automates the creation of the API-MDE bridge. Our proposal includes a complete prototype of a toolkit focused on Java APIs, although an adaptation of the approach to deal with APIs for other statically-typed object-oriented languages (such as C sharp) could be easily implemented.

Security Information Discovery Most companies information systems are composed by heterogeneous components responsible of hosting, creating or manipulating critical information for the day-to-day operation of the company. Securing this information is therefore one of their main concerns, more particularly specifying Access Control (AC) policies. However, the task of implementing an AC security policy (sometimes relying on several mechanisms) remains complex and error prone as it requires knowing low level and vendor-specific facilities. In this context, discovering and understanding which security policies are actually being enforced by the Information System (IS) becomes critical. Thus, the main challenge consists in bridging the gap between the vendor-dependent security features and a higher-level representation. This representation has to express the policies by abstracting from the specificities of the system components, allowing security experts to better understand the policy and to implement all related evolution, refactoring and manipulation operations in a reusable way. As a first result, in [28] a method to extract AC policies from firewall configuration files is proposed.

Business Rules Discovery In order to react to the ever-changing market, every organization needs to periodically reevaluate and evolve its company policies. These policies must be enforced by its Information System (IS) by means of a set of so-called business rules that drive the system behavior and data. Clearly, policies and rules must be aligned at all times but unfortunately this is a challenging task. In most ISs, the implementation of business rules is scattered among the code so appropriate techniques must be provided for the discovery and evolution of changing business rules. In [24], we describe a MDRE framework aiming at extracting business rules out of Java source code. The use of modeling techniques facilitate the representation of the rules at a higher-abstraction level which enables stakeholders to understand and manipulate them more easily.

Software Modernization Software modernization processes usually follow the well-known horse-shoe model, which provides a framework to integrate different abstraction levels and reverse engineering tools. The Architecture-Driven Modernization (ADM) is an OMG's initiative which aims at defining and standardizing techniques, methods and tools for software modernization. It incorporates the horse-shoe framework as its reference model and uses MDE techniques as the implementation foundation. Since ADM proposes applying the modernization process at the most abstract level, we believe that, to some extent, the ADM initiative has misinterpreted the original horse-shoe model [34].

Legacy Data Federation The fast evolution of technologies (SOA, Cloud, mobile environments), ISs complexity and the growing need for agility require to be able to represent information systems as a whole. In this context, Enterprise Architecture (EA) approaches intend to address all the systems dimensions: software components, associated physical resources, relationships with the companies requirements and business processes, implied actors/roles/structures, etc. Within the TEAP FUI project (cf. corresponding section), we have started studying the reverse engineering capabilities required when dealing with such high-level views of an IS. More particularly, the focus has been put on features for allowing federating the relevant data coming from different existing sources, as well as for integrating them efficiently. To this intent, a prototype is currently being developed based on several technologies from the team (e.g. Virtual EMF, ATL, MoDisco).

6.6. Empirical software modeling

A new line started this year was the evaluation of how software modeling techniques (and in general software engineering methods) are used in practice. As the first area of study, we have focused on how software architects deal with non-functional requirements. Based on a set of interviews with software architects, we have analyzed whether all the languages, patterns and methodologies proposed by researchers have had any impact on the way software architect choose the best architecture for a given system. Results of the study can be read in these publications [18] [11].

AVALON Team

6. New Results

6.1. HPC Component Model

Participants: Zhengxiong Hou, Vincent Pichon, Christian Pérez.

6.1.1. L2C: A Low Level Component Model

We have proposed a low level component model (L^{2}C) that supports directly native connectors for typical scenarios of high performance computing, such as MPI, shared memory and method invocation [10]. We have applied it to a typical example of stencil computation, i.e., a 2-D Jacobi application with domain decomposition. The experimental results have shown that L^{2}C can achieve the equivalent performance as native implementations, while gaining benefits such as performance portability on the basis of the software component model.

6.1.2. Auto-tuning of Stencil Based Applications

We started modeling the performance of stencil applications on multi-core clusters. We focused in particular on a 2D Jacobi benchmark application and the NEMO application as well as memory bandwidth performance. We derived a tuning approach including data partitioning within one node, the selection of the number of threads within a multi-core node, a data partitioning for multi nodes, and the number of nodes for a multi-core cluster. This model is based on a set of experiments on machines of GRID'5000 and on Curie and Juqueen supercomputers. A paper presenting these results is in preparation.

6.2. Cooperative Resource Managers

Participants: Eddy Caron, Cristian Klein, Christian Pérez, Noua Toukourou.

6.2.1. Integration of SALOME with CooRM

We have continued the validation works of the CooRM RMS architecture [52]. To this end, we focused on the SALOME numerical simulation platform developed and used jointly by EDF and CEA. In 2012, we have mostly started the integration of CooRMv1 concepts in SALOME. CooRMv1 targets moldable applications and allows them to efficiently employ their custom resource selection algorithms. We have done the necessary changes in SALOME, thus obtaining a working prototype implementation. Thanks to this, SALOME applications could be published with a custom launcher (implementing a resource selection algorithm) so as to transparently launch applications efficiently, instead of having to leave this burden to the user.

6.2.2. A Distributed Resource Management Architecture for Moldable Applications

In 2011, we have proposed CooRMv1 [52], a centralized RMS architecture to efficiently support moldable applications. Having a centralized architecture is however undesirable for geographically-distributed resources such as Grids or multiple Clouds. For example, if there is a network failure, some users will not be able to access any resources, not even those that are located on their side of the bisection.

To this end, we extended CooRMv1 and proposed a distributed version of it, distCooRM, in collaboration with the Myriads team. It allows moldable applications to efficiently co-allocated resource managed by independent agents. Simulation results show that the approach is feasible and scales well for a reasonable number of applications. In other words, it presents good strong scalability, but not weak scalability, which we intend to address in future work.

6.2.3. A Resource Management Architecture for Fair Scheduling of Optional Computations

In collaboration with two teams from IRIT, we have identified a use-case that is currently badly supported. Some applications, such as Monte-Carlo simulations, contain optional computations: These are not critical, but completing them would improve the results. When executing these application on HPC resources, most resource managers, such as batch schedulers, require the user to submit a predefined number of computing requests. If the user submits too many requests, the platform might become overloaded, whereas if the user submits too few requests, then resources might be left idle.

To solve this issue, we proposed a resource management architecture, called DIET-ethic [42], which auto-tunes the number of optional requests. It improves user happiness, fairness and the number of completed requests, when compared to a system which does not support optional computations.

6.3. Large-Scale Data Management and Processing

Participants: José Saray, Bing Tang, Gilles Fedak, Anthony Simonet.

6.3.1. Data Management on Hybrid Distributed Infrastructure

The BITDEW framework addresses the issue of how to design a programmable environment for automatic and transparent data management on Grids, Clouds and Desktop Grids. BITDEW relies on a specific set of meta-data to drive key data management operations, namely life cycle, distribution, placement, replication and fault-tolerance with a high level of abstraction.

In collaboration with Mohamed Labidi, University of Sfax (Tunisia), we have developed a data-aware and parallel version of Magik, an application for Arabic writing recognition using the BITDEW middleware. We are targeting digital libraries, which require distributed computing infrastructure to store the large number of digitalized books as raw images and at the same time to perform automatic processing of these documents such as OCR, translation, indexing, searching, etc. [20].

In 2012, we have also surveyed P2P strategies (replication, erasure code, replica repair, hybrid storage), which provide reliable and durable storage on top of hybrid distributed infrastructures composed of volatile and stable storage. Following these simulation studies, we are implementing a prototype of the Amazon S3 storage on top of BitDew, which will provide reliable storage by using both Desktop free disk space and volunteered remote Cloud storage [25].

6.3.2. MapReduce Programing Model for Desktop Grid

MapReduce is an emerging programming model for data-intense applications proposed by Google, which has recently attracted a lot of attention. MapReduce borrows from functional programming, where programmer defines Map and Reduce tasks executed on large sets of distributed data. In 2010, we developed an implementation of the MapReduce programming model based on the BitDew middleware. Our prototype features several optimizations which make our approach suitable for large scale and loosely connected Internet Desktop Grid: massive fault tolerance, replica management, barriers-free execution, latency-hiding optimization as well as distributed result checking. We have presented performance evaluations of the prototype both against micro-benchmarks and real MapReduce applications. The scalability test achieved linear speedup on the classical WordCount benchmark. Several scenarios involving lagger hosts and host crashes demonstrated that the prototype is able to cope with an experimental context similar to real-world Internet [9].

In collaboration with the Huazhong University of Science & Technology (China), we have developed an emulation framework to assess MapReduce on Internet Desktop Grid. We have made extensive comparison on BitDew-MapReduce and Hadoop using GRID'5000 which show that our approach has all the properties desirable to cope with an Internet deployment, whereas Hadoop fails on several tests [22].

We have published a joint work in collaboration with Virginia Tech (USA), which is a presentation of two alternative implementations of MapReduce for Desktop Grids : Moon and Bitdew [37].

6.4. Computing on Hybrid Distributed Infrastructure

Participants: Simon Delamare, Gilles Fedak, José Saray, Anthony Simonet.

6.4.1. SpeQuloS: Providing Quality-of-Service to Desktop Grids using Cloud resources

EDGI is an FP7 European project, following the successful FP7 EDGeS project, whose goal is to build a Grid infrastructure composed of "Desktop Grids", such as BOINC or XtremWeb, where computing resources are provided by Internet volunteers, and "Service Grids", where computing resources are provided by institutional Grid such as EGI, gLite, Unicore and "Clouds systems" such as OpenNebula and Eucalyptus, where resources are provided on-demand. The goal of the EDGI project is to provide an infrastructure where Service Grids are extended with public and institutional Desktop Grids and Clouds.

The main limitation with the current infrastructure is that it cannot give any QoS support for applications running in the Desktop Grid (DG) part of the infrastructure. For example, a public DG system enables clients to return work-unit results in the range of weeks. Although there are EGI applications (e.g., the fusion community's applications) that can tolerate such a long latency most of the user communities want much shorter deadlines.

In 2011, we have developed the SpeQuloS middleware to solve this critical problem. Providing QoS features even in Service Grids is hard and not solved yet satisfactorily. It is even more difficult in an environment where there are no guaranteed resources. In DG systems, resources can leave the system at any time for a long time or forever even after taking several work-units with the promise of computing them. Our approach is based on the extension of DG systems with Cloud resources. For such critical work-units the SpeQuloS system is able to dynamically deploy fast and trustable clients from some Clouds that are available to support the EDGI DG systems. It takes the right decision about assigning the necessary number of trusted clients and Cloud clients for the QoS applications. In 2012, we have conducted extensive simulations to evaluate various strategies of Cloud resources provisioning. Results show that SpeQuloS improve the QoS of BoTs on three aspects: it reduces the makespan by removing the tail effect, it improves the execution stability and it allows to accurately predicts the BoT completion time [14], [21], [35]. The software have now been delivered to the partners and run in production in the European Desktop Grid Infrastructure.

6.4.2. Scheduling on Hybrid Distributed Computing Infrastructures

In collaboration with the Mircea Moca, from the Babes-Bolyai University of Cluj-Napoca (Romania), we have investigated new scheduling algorithms for pull-based scheduler, which relies on Promethee method. We have shown that these heuristics perform efficiently on three different kinds of infrastructures, namely Grids, Clouds and Desktop Grids [23].

6.5. Energy Efficiency in Large Scale Systems

Participants: Ghislain Landry Tsafack, Mohammed El Mehdi Diouri, Olivier Glück, Laurent Lefevre.

6.5.1. Energy Efficiency in HPC Systems

Modern high performance computing subsystems (HPC) – including processor, network, memory, and I/O — are provided with power management mechanisms. These include dynamic speed scaling and dynamic resource sleeping. Understanding the behavioral patterns of high performance computing systems at runtime can lead to a multitude of optimization opportunities including controlling and limiting their energy usage. We have proposed a general purpose methodology for optimizing energy performance of HPC systems considering processor, disk and network. We have relied on the concept of execution vector along with a partial phase recognition technique for on-the-fly dynamic management without any a priori knowledge of the workload. We have demonstrated the effectiveness of our management policy under two real-life workloads. Experimental results have shown that our management policy in comparison with baseline unmanaged execution saves up to 24% of energy with less than 4% performance overhead for our real-life workloads [28], [27], [26]. This work is done under the Large Scale Initiative Hemera project (Joint PhD between Avalon and IRIT (Toulouse) with J.-M. Pierson, P. Stolf and G. Da Costa).

6.5.2. Energy Considerations in Checkpointing and Fault Tolerance Protocols

Two key points should be taken into account in future exascale systems: fault tolerance and energy consumption. To address these challenges, we evaluated checkpointing and existing fault tolerance protocols from an energy point of view. We measured on a real testbed the power consumption of the main atomic operations found in these protocols: checkpointing, message logging and coordination. The results [16], [51] show that process coordination and RAM logging consume more power than checkpointing and HDD logging. However, the results we presented in Joules per Bytes for I/O operations, emphasize that checkpointing and HDD logging consume more energy than RAM logging because of the logging duration which is much more higher on HDD than on RAM . We have also shown that for identical nodes performing the same operation, the extra power cost due to this operation is the same. In general, we have learned that the power consumption of a node during a given operation remains constant during this operation. The power consumption of such a node is equal to its idle power consumption to which we add the extra power consumption due to the operation it is performing. Finally, we proposed to consider energy consumption as a criterion for the choice of fault tolerance protocols. In terms of energy consumption, we should promote message logging for applications exchanging small volumes of data and coordination for applications involving few processes. This work is a joint work with F. Cappello (Inria-UIUC-NCSA Joint Laboratory for Petascale Computing).

6.5.3. Towards a Smart and Energy-Aware Service-Oriented Manager for Extreme-Scale Applications

To address the issue of energy efficiency for exascale supercomputers, we proposed a smart and energy-aware service-oriented manager for exascale applications: SEASOMES [17]. This framework aggregates the various energy-efficient solutions to "consume less" energy and to "consume better". It involves both internal and external interactions with the various actors interfering directly or indirectly with the supercomputer. On the one hand, we recommended a more fine-grained collaboration between application and hardware resources in order to reduce energy consumption and provide sustainable exascale services. On the other hand, we suggested a cooperation between the user, the administrator, the resource manager and the energy supplier for the purpose of "consuming better".

6.6. Green-IT Innovation Analysis

Participant: Laurent Lefevre.

Green IT has recently appeared as a mandatory approach to take into account of energy efficiency in Information Technology. This research investigates the Green IT area and its opportunities for innovation. Main motivations for Green IT have been analyzed and we have proposed new definition of Green IT including social, environmental and economic concerns. We have proposed a new model of a virtuous circle that appears in Green IT: while Green IT has its own motivations, resulting research feeds other research field in a virtuous circle. Innovation in this particular sector paves the way for further innovation by means of original research not foreseen at first thoughts.

This analysis is joint work with IRIT (Toulouse - C. Herozog, J.-M. Pierson) [19].

6.7. Workflow Scheduling

Participants: Eddy Caron, Frédéric Desprez, Cristian Klein, Vincent Lanore, Sylvain Gault, Christian Pérez, Adrian Muresan, Frédéric Suter.

6.7.1. High-Level Waste Application Scheduling

Brought forward by EDF, a partner in the ANR COOP project, High-Level Waste is a multi-level application: It is composed of many moldable tasks, part of which are initially known. Some of these tasks may, with a certain probability, launch other tasks, which usually take longer. We have proposed several scheduling algorithms to optimize the performance of such applications, which are little studied in current literature. Experiments with simulations showed that considerable gains can be made, not only in terms of performance, but also performance portability. This work will be published in 2013 [31].

6.7.2. Elastic Scheduling for Functional Workflows

As a recent research direction we have focused on the development of an allocation strategy for budget-constrained workflow applications that target IaaS Cloud platforms. The workflow abstraction is very common amongst scientific applications. It is easy to find examples in any field from bioinformatics to geography. The reasons for the proliferation of workflow applications in science are various, from the building of applications on top of legacy code to modeling of applications that have an inherent workflow structure. The first workflow applications were composed of sequential tasks, but as computational units became more and more parallel, workflow applications have also evolved and are now formed of parallel tasks and, occasionally, parallel moldable tasks. The classic DAG structure of workflow applications has also changed as some applications need to perform refinement iteration, creating loop-like constructs.

We have considered a general model of workflow applications that permit non-deterministic transitions. We have elaborated two budget-constrained allocation strategies for this type of workflow. The problem is a bi-criteria optimization problem as we are optimizing both budget and workflow makespan [12].

For a practical validation of the work, we are currently working on the implementation of the budget-constrained scheduler as part of the Nimbus open source cloud platform. This is being tested with a cosmological simulation workflow application called *Ramses* (see Section 4.4). This is a parallel MPI application that, as part of this work, has been ported for execution on dynamic virtual platforms. This work has been done in the form of a two month internship at the Argonne National Laboratory, USA, under the guidance of Kate Keahey and has been accepted for poster presentation in the XSEDE 2012 conference.

6.7.3. Self-Healing of Operational Workflow Incidents on Distributed Computing Infrastructures

Distributed computing infrastructures are commonly used through scientific gateways, but operating these gateways requires important human intervention to handle operational incidents. We have designed a self-healing process that quantifies incident degrees of workflow activities from metrics measuring long-tail effect, application efficiency, data transfer issues, and site-specific problems. These metrics are simple enough to be computed online and they make little assumptions on the application or resource characteristics. From their degree, incidents are classified in levels and associated to sets of healing actions that are selected based on association rules modeling correlations between incident levels. We specifically study the long-tail effect issue, and propose a new algorithm to control task replication. The healing process is parametrized on real application traces acquired in production on the European Grid Infrastructure. Experimental results obtained in the Virtual Imaging Platform show that the proposed method speeds up execution up to a factor of 4, consumes up to 26% less resource time than a control execution and properly detects unrecoverable errors.

This work is done in collaboration with Tristan Glatard and Rafael Ferreira Da Silva from CREATIS (UMR5220).

6.7.4. Scheduling for MapReduce Based Applications

We have worked on scheduling algorithms for MapReduce applications in Grids and Clouds as we aim at providing resource-efficient and time-efficient scheduling algorithms. This work is mainly done within the scope of the Map-Reduce ANR project.

A deliverable presenting the heuristics for scheduling data transfers derived from a previous work by Berlinska and Drozdowsky has been written [50]. A section of a collaborative paper has been written and the paper has been presented at the ICA CON conference [9], [4]. The results of the aforementioned heuristics that has been previously implemented in a visualization / simulation tool, has been summarized in a paper accepted for RenPar. Moreover, these algorithms and heuristics have been implemented in the MapReduce framework HoMR.

6.8. Performance Evaluation and Modeling

Participants: Eddy Caron, Frédéric Desprez, Matthieu Imbert, Georges Markomanolis, Jonathan Rouzaud-Cornabas, Frédéric Suter.

6.8.1. Time-Independent Log Format

Simulation is a popular approach to obtain objective performance indicators of platforms that are not at one's disposal. It may for example help the dimensioning of compute clusters in large computing centers. In many cases, the execution of a distributed application does not behave as expected, it is thus necessary to understand what causes this strange behavior. Simulation provides the possibility to reproduce experiments under similar conditions. This is a suitable method for experimental validation of a parallel or distributed application.

The tracing instrumentation of a profiling tool is the ability to save all the information about the execution of an application at run-time. Every scientific application executed computed instructions. The originality of our approach is that we measure the completed instructions of the application and not its execution time. This means that if a distributed application is executed on N cores and we execute it again by mapping two processes per core then we need N/2 cores and more time for the execution time of the application. An execution trace of an instrumented application can be transformed into a corresponding list of actions. These actions can then be simulated by SimGrid. Moreover the SimGrid execution traces will contain almost the same data because the only change is the use of half cores but the same number of processes. This does not affect the number of the completed instructions so the simulation time does not get increased because of the overhead. The GRID'5000 platform is used for this work and the NAS Parallel Benchmarks are used to measure the performance of the clusters.

Our main contribution is to propose of a new execution log format that is time-independent. This means that we decouple the acquisition of the traces from the replay. Furthermore we implemented a trace replay tool which relies on top of fast, scalable and validated simulation kernel of SimGrid. We proved that this framework applies for some of the NAS Parallel Benchmarks and we can predict their performance with a good accuracy. Moreover we improved the accuracy of the performance's prediction by applying different instrumentation configurations according to the requirements of our framework. Some performance issues of the executed benchmarks were taken under consideration for more accurate predictions. Also the simulator was reimplemented in order to have more accurate results and take advantage of the last SimGrid's simulation techniques. Finally we did a survey on many different tracing tools with regards to the requirements of our methodology which includes all the latest provided tools from the community. For the extreme cases where we used many nodes by mapping a lot of processes per core, some issues were indicated that we are trying to solve in order to be able to apply our methodology with less overhead. Also we plan to predict the performance of more benchmarks.

6.8.2. Dynamic Network Forecasting

In distributed systems the knowledge of the network is mandatory to know the available connections and their performance. Indeed, to be able to efficiently schedule network transfers on computing platforms such as clusters, grids or clouds, accurate and timely predictions of network transfers completion times are needed. We designed a new metrology and performance prediction framework called Pilgrim which offers a service predicting the completion times of current and concurrent TCP transfers. This service uses SimGrid to simulate the network transfers. Ongoing work is to obtain experimental results comparing the predictions obtained from Pilgrim to the real transfer completion times.

6.8.3. Amazon EC2 simulation

During this year, we have developed an extension of SimGrid to simulate multi-platforms Clouds: SimGrid Cloud Broker (SGCB). It simulates the suite of services provided by Amazon AWS: EC2 for virtual machines, S3 for key-value storage and EBS for block storage. SGCB allows to easily evaluate different resource selection policy but also to simulate an entire application running on a set of resources that come from multiple Clouds. As the billing mechanism is a crucial feature of the Clouds, SGCB is able to simulate it. For this, we extended SimGrid in order to do the accounting of all virtual resources used. With this accounting, we are able to simulate the process of billing as Amazon does it. We are working to increase the accuracy of our performance models, and therefore the validity of the results for different use cases.

6.9. Cloud Resource Management

Participants: Eddy Caron, Frédéric Desprez, Arnaud Lefray, Jonathan Rouzaud-Cornabas, Julien Carpentier, Jean-Patrick Gelas, Laurent Lefevre, Maxime Morel, Olivier Mornard, Francois Rossigneux.

6.9.1. Resource Provisioning for Federations of Clouds

Since the visit of Jose Luis Lucas Simarro, we have established a collaboration with the Distributed Systems Architecture Research Group at Complutense University of Madrid (Spain) on resource brokering strategies for multiples Clouds. The purpose is to design new strategies that are able to migrate services from a Cloud to another one. VM migration is done to save money when the price of running a given VM change. Indeed, in modern Clouds such as Amazon EC2, Spot Instances have dynamic prices that change based on the law of supply and demand. Most of the current solutions only take into account the cost of computation when migrating services between Clouds. However, when a service is migrated, we need to pay network traffic between the two Clouds and the storage of the Virtual Machine image in both Clouds during the migration. We are studying trough simulations different resource selection algorithms that take into account the cost of all resources: compute, storage, and network.

6.9.2. Energy Efficient Clouds

Within the projects CompatibleOne (Open Source Cloud Broker) and XLcloud (Energy Efficiency in OpenStack based clouds), we explore the design of energy aware and energy efficient cloud infrastructures. Monitoring of physical and virtual resources is injected into cloud frameworks. Systems based on such metrics are designed in order to benefit from energy usage knowledge in virtual machines mapping and precise accounting [13].

6.9.3. User Isolation

Inter-VM and virtual network isolation is weak in terms of both security and performance. Accordingly, it can not guarantee performance, security and privacy requirements. This is a serious issue as most of clouds are multi-tenant and users do not trust each other. By improving the resource allocation process, we show how these issue can be solved and thus the overall security of the clouds improved. Moreover, we show how a Cloud Service Provider (CSP) can let the users express their security requirements. We show that isolation requirements have a cost for the Cloud Service Providers but they can bill requirements as an additional service. By doing so, they will have a new resource of income and the users trust in their platforms will increase as they can express security requirements.

6.9.4. Cloud Security

Mandatory Access Control is really poorly supported by Cloud environments. Our work proposes extensions of the OpenNebula Cloud in order to provide an advanced MAC protection of the virtual machines hosted by the different nodes of the Cloud. Thus, unique SELinx security labels are associated with the virtual machines and their resources. The instantiations and migrations of the virtual machines maintain those unique security labels. Moreover, PIGA-Virt provides a unified way to control the information flows within a virtual machine but also between multiple virtual machines. SELinux controls the direct flows. PIGA-Virt adds advanced controls. Thus, a PIGA protection rule can control several direct and indirect flows. The benchmarks of PIGA-Virt show that our Trusted OpenNebula Cloud is efficient regarding the quality of the protection.

This work is done in collaboration with Christian Toinard from LIFO/ENSI de Bourges.

6.10. Virtualizing Home Gateways at Large Scale

Participants: Jean-Patrick Gelas, Laurent Lefevre.

About 80-90% of the energy in today's wireline networks is consumed in the access network, with about 10 W per user being dissipated mostly by the customer premises equipment (CPE). Home gateway is a popular equipment deployed at the end of networks and supporting a set of heterogeneous services (from network to multimedia services). These gateways are difficult to manage for network operators and consume a lot of energy. This research explores the possibility to reduce the complexity of such equipment by moving services to some external dedicated and shared equipments. When combined to quasi passive CPE, this approach can reduce the energy consumption of wired networks infrastructures. This research is done within the GreenTouch initiative which aims to increase network energy efficiency by a factor of 1000 from current levels by 2015.

This work is done with collaboration with Addis Abeba University (Ethiopia) (M. Mulugeta and T. Assefa) [18].

6.11. Self-Adaptive Deployment

Participants: Eddy Caron, Maurice-Djibril Faye, Jonathan Rouzaud-Cornabas.

Software systems are increasingly expected to be self-adaptive. Such software systems have the capability to autonomously modify their behavior at run-time in response to changes in their environment. This capability may be included in the software systems at design time or later by external mechanisms. Therefore, along their development process multiple adaptation concerns must be considered, such as the response to changes in the utilization patterns, the need for alternative algorithms for implementing a function, or the diversity of the infrastructure. We have designed an architecture which aims to add self-adaptive capabilities to an existing middleware so that its deployment becomes self-adaptive. The framework uses external mechanisms for that purpose since this capability was not a native feature.

CEPAGE Project-Team

6. New Results

6.1. Resource allocation and Scheduling

6.1.1. Divisible Load Scheduling

Participants: Olivier Beaumont, Nicolas Bonichon, Lionel Eyraud-Dubois.

Malleable tasks are jobs that can be scheduled with preemptions on a varying number of resources. In [22], we focus on the special case of work-preserving malleable tasks, for which the area of the allocated resources does not depend on the allocation and is equal to the sequential processing time. Moreover, we assume that the number of resources allocated to each task at each time instant is bounded. We consider both the clairvoyant and non-clairvoyant cases, and we focus on minimizing the weighted sum of completion times. In the weighted non-clairvoyant case, we propose an approximation algorithm whose ratio (2) is the same as in the unweighted non-clairvoyant case. In the clairvoyant case, we provide a normal form for the schedule of such malleable tasks, and prove that any valid schedule can be turned into this normal form, based only on the completion times of the tasks. We show that in these normal form schedules, the number of preemptions per task is bounded by 3 on average. At last, we analyze the performance of greedy schedules, and prove that optimal schedules are greedy for a special case of homogeneous instances. We conjecture that there exists an optimal greedy schedule for all instances, which would greatly simplify the study of this problem. Finally, we explore the complexity of the problem restricted to homogeneous instances, which is still open despite its very simple expression. (Joint work with Loris Marchal from ENS Lyon)

6.1.2. Scheduling for Distributed Continuous Integration

Participants: Olivier Beaumont, Nicolas Bonichon, Ludovic Courtès.

In [21], we consider the problem of schedul- ing a special kind of mixed data-parallel applications arising in the context of continuous integration. Continuous integration (CI) is a software engineering technique, which consists in re- building and testing interdependent software components as soon as developers modify them. The CI tool is able to provide quick feedback to the developers, which allows them to fix the bug soon after it has been introduced. The CI process can be described as a DAG where nodes represent package build tasks, and edges represent dependencies among these packages; build tasks themselves can in turn be run in parallel. Thus, CI can be viewed as a mixed data-parallel application. A crucial point for a successful CI process is its ability to provide quick feedback. Thus, makespan minimization is the main goal. Our contribution is twofold. First, we provide and analyze a large dataset corresponding to a build DAG. Second, we compare the performance of several scheduling heuristics on this dataset.

6.1.3. Resource Allocation in Clouds

Participants: Olivier Beaumont, Lionel Eyraud-Dubois, Hejer Rejeb.

In [14], we consider the problem of assigning a set of clients with demands to a set of servers with capacities and degree constraints. The goal is to find an allocation such that the number of clients assigned to a server is smaller than the server's degree and their overall demand is smaller than the server's capacity, while maximizing the overall throughput. This problem has several natural applications in the context of independent tasks scheduling or virtual machines allocation. We consider both the *offline* (when clients are known beforehand) and the *online* (when clients can join and leave the system at any time) versions of the problem. We first show that the degree constraint on the maximal number of clients that a server can handle is realistic in many contexts. Then, our main contribution is to prove that even if it makes the allocation problem more difficult (NP-Complete), a very small additive resource augmentation on the servers degree is enough to find in polynomial time a solution that achieves at least the optimal throughput. After a set of theoretical results on the complexity of the offline and online versions of the problem, we propose several other greedy heuristics to solve the online problem and we compare the *performance* (in terms of throughput) and the *cost* (in terms of disconnections and reconnections) of all proposed algorithms through a set of extensive simulation results. (Joint work with Christopher Thraves-Caros, University of Madrid)

6.1.4. Non Linear Divisible Load Scheduling

Participants: Olivier Beaumont, Hubert Larchevêque.

Divisible Load Theory (DLT) has received a lot of attention in the past decade. A divisible load is a perfect parallel task, that can be split arbitrarily and executed in parallel on a set of possibly heterogeneous resources. The success of DLT is strongly related to the existence of many optimal resource allocation and scheduling algorithms, what strongly differs from general scheduling theory. Moreover, recently, close relationships have been underlined between DLT, that provides a fruitful theoretical framework for scheduling jobs on heterogeneous platforms, and MapReduce, that provides a simple and efficient programming framework to deploy applications on large scale distributed platforms. The success of both have suggested to extend their framework to non-linear complexity tasks. In [24], we show that both DLT and MapReduce are better suited to workloads with linear complexity. In particular, we prove that divisible load theory cannot directly be applied to quadratic workloads, such as it has been proposed recently. We precisely state the limits for classical DLT studies and we review and propose solutions based on a careful preparation of the dataset and clever data partitioning algorithms. In particular, through simulations, we show the possible impact of this approach on the volume of communications generated by MapReduce, in the context of Matrix Multiplication and Outer Product algorithms. (Joint work with Loris Marchal from ENS Lyon)

6.1.5. Reliable Service Allocation in Clouds

Participants: Olivier Beaumont, Lionel Eyraud-Dubois, Hubert Larchevêque.

In [23], we consider several reliability problems that arise when allocating applications to processing resources in a Cloud computing platform. More specifically, we assume on the one hand that each computing resource is associated to a capacity constraint and to a probability of failure. On the other hand, we assume that each service runs as a set of independent instances of identical Virtual Machines, and that the Service Level Agreement between the Cloud provider and the client states that a minimal number of instances of the service should run with a given probability. In this context, given the capacity and failure probabilities of the machines, and the capacity and reliability demands of the services, the question for the cloud provider is to find an allocation of the instances of the services (possibly using replication) onto machines satisfying all types of constraints during a given time period. The goal of this work is to assess the impact of the reliability constraint on the complexity of resource allocation problems. We consider several variants of this problem, depending on the number of services and whether their reliability demand is individual or global. We prove several fundamental complexity results (#P' and NP-completeness results) and we provide several optimal and approximation algorithms. In particular, we prove that a basic randomized allocation algorithm, that is easy to implement, provides optimal or quasi-optimal results in several contexts, and we show through simulations that it also achieves very good results in more general settings.

6.1.6. Optimizing Resource allocation while handling SLA violations in Cloud Computing platforms

Participants: Lionel Eyraud-Dubois, Hubert Larchevêque.

In [29], we study a resource allocation problem in the context of Cloud Computing, where a set of Virtual Machines (VM) has to be placed on a set of Physical Machines (PM). Each VM has a given demand (e.g. CPU demand), and each PM has a capacity. However, each VM only uses a fraction of its demand. The aim is to exploit the difference between the demand of the VM and its real utilization of the resources, to exploit the capacities of the PMs as much as possible. Moreover, the real consumption of the VMs can change over time (while staying under its original demand), implying sometimes expensive "SLA violations", corresponding to some VM's consumption not satisfied because of overloaded PMs. Thus, while optimizing the global resource utilization of the PMs, it is necessary to ensure that at any moment a VM's need evolves, a few number of migrations (moving a VM from PM to PM) is sufficient to find a new configuration in which all the VMs' consumptions are satisfied. We modelize this problem using a fully dynamic bin packing approach and we present an algorithm ensuring a global utilization of the resources of 66%. Moreover, each time a PM is overloaded at most one migration is necessary to fall back in a configuration with no overloaded PM, and only

3 different PMs are concerned by required migrations that may occur to keep the global resource utilization correct. This allows the platform to be highly resilient to a great number of changes.

6.2. Compact Routing

6.2.1. Compact routing with forbidden-set in planar graphs

Participant: Cyril Gavoille.

In [20], we consider fully dynamic $(1+\varepsilon)$ distance oracles and $(1+\varepsilon)$ forbidden-set labeling schemes for planar graphs. For a given n-vertex planar graph G with edge weights drawn from $[1, M]$ and parameter $\varepsilon > 0$, our forbidden-set labeling scheme uses labels of length $\lambda = O(\varepsilon^{-1} \log^2 n \log{(nM)} \cdot \max{logn})$. Given the labels of two vertices s and t and of a set F of faulty vertices/edges, our scheme approximates the distance between s and t in $G \setminus F$ with stretch $(1+\varepsilon)$, in $O(|F|^2\lambda)$ time.

We then present a general method to transform $(1+\varepsilon)$ forbidden-set labeling schemas into a fully dynamic $(1+\varepsilon)$ distance oracle. Our fully dynamic $(1+\varepsilon)$ distance oracle is of size $O(n \log n \cdot \max{logn})$ and has $\widetilde{O}(n^{1/2})$ query and update time, both the query and the update time are worst case. This improves on the best previously known $(1+\varepsilon)$ dynamic distance oracle for planar graphs, which has worst case query time $\widetilde{O}(n^{2/3})$ and amortized update time of $\widetilde{O}(n^{2/3})$.

Our $(1+\varepsilon)$ forbidden-set labeling scheme can also be extended into a forbidden-set labeled routing scheme with stretch $(1+\varepsilon)$.

6.2.2. Planar Spanner of geometric graphs

Participants: Nicolas Bonichon, Cyril Gavoille, Nicolas Hanusse.

In [26], we determine the stretch factor of L_1-Delaunay and L_∞-Delaunay triangulations, and we show that this stretch is $\sqrt{4+2\sqrt{2}} \approx 2.61$. Between any two points x, y of such triangulations, we construct a path whose length is no more than $\sqrt{4+2\sqrt{2}}$ times the Euclidean distance between x and y, and this bound is best possible. This definitively improves the 25-year old bound of $\sqrt{10}$ by Chew (SoCG '86).

To the best of our knowledge, this is the first time the stretch factor of the well-studied L_p-Delaunay triangulations, for any real $p \geq 1$, is determined exactly.

6.3. Mobile Agents

6.3.1. More efficient periodic traversal in anonymous undirected graphs

Participants: David Ilcinkas, Ralf Klasing.

In [15], we consider the problem of *periodic graph exploration* in which a mobile entity with constant memory, *an agent*, has to visit all n nodes of an input simple, connected, undirected graph in a periodic manner. Graphs are assumed to be anonymous, that is, nodes are unlabeled. While visiting a node, the agent may distinguish between the edges incident to it; for each node v, the endpoints of the edges incident to v are uniquely identified by different integer labels called *port numbers*. We are interested in algorithms for assigning the port numbers together with traversal algorithms for agents using these port numbers to obtain short traversal periods.

Periodic graph exploration is unsolvable if the port numbers are set arbitrarily; see Budach (1978). However, surprisingly small periods can be achieved by carefully assigning the port numbers. Dobrev *et al.* (2005) described an algorithm for assigning port numbers and an oblivious agent (i.e., an agent with no memory) using it, such that the agent explores any graph with n nodes within the period $10n$. When the agent has access to a constant number of memory bits, the optimal length of the period was proved in Gasieniec *et al.* (2008) to be no more than $3.75n - 2$ (using a different assignment of the port numbers and a different traversal algorithm). In our work, we improve both these bounds. More precisely, we show how to achieve a period length of at most $(4 + \frac{1}{3})n - 4$ for oblivious agents and a period length of at most $3.5n - 2$ for agents with constant memory. To obtain our results, we introduce a new, fast graph decomposition technique called a *three-layer partition* that may also be useful for solving other graph problems in the future. Finally, we present the first non-trivial lower bound, $2.8n - 2$, on the period length for the oblivious case.

6.3.2. Gathering of Robots on Anonymous Grids without Multiplicity Detection

Participant: Ralf Klasing.

In [28] , we study the gathering problem on grid networks. A team of robots placed at different nodes of a grid have to meet at some node and remain there. Robots operate in Look-Compute-Move cycles; in one cycle, a robot perceives the current configuration in terms of occupied nodes (Look), decides whether to move towards one of its neighbors (Compute), and in the positive case makes the computed move instantaneously (Move). Cycles are performed asynchronously for each robot. The problem has been deeply studied for the case of ring networks. However, the known techniques used on rings cannot be directly extended to grids. Moreover, on rings, another assumption concerning the so-called *multiplicity detection* capability was required in order to accomplish the gathering task. That is, a robot is able to detect during its Look operation whether a node is empty, or occupied by one robot, or occupied by an undefined number of robots greater than one.

In our work, we provide a full characterization about gatherable configurations for grids. In particular, we show that in this case, the multiplicity detection is not required. Very interestingly, sometimes the problem appears trivial, as it is for the case of grids with both odd sides, while sometimes the involved techniques require new insights with respect to the well-studied ring case. Moreover, our results reveal the importance of a structure like the grid that allows to overcome the multiplicity detection with respect to the ring case.

CIDRE Project-Team

6. New Results

6.1. Intrusion Detection

6.1.1. Intrusion Detection based on an Analysis of the Flow Control

In 2012, we strengthened our research efforts around intrusion detection parameterized by a security policy.

In [22] we formally study information flows that occur during the executions of a system implementing a classical access control mechanism. More precisely, we detail how the generic access control model we proposed defines two sets of illegal information flows: the first set corresponds to the flows resulting from the accesses authorized by the access control policy while the second set corresponds to the information flow policy deduced from the access control policy interpretation. We show that these two sets may coincide for some policies and we propose a mechanism dedicated to illegal information flow detection that can be useful in other cases. Finally, we describe a real implementation for the Linux operating system.

In [38], we extended our previous illegal information flow detector to track network exchanges. A confidentiality policy is defined by labeling sensitive information and defining which information may leave the local system through network exchanges. Furthermore, per application profiles can be defined to restrict the sets of information each application may access and/or send through the network. An example application of this extension in the context of a compromised web browser showed that our implementation can detect a confidentiality violation when the browser attempts to leak private information to a remote host over the network.

In [30], we adapted our detection model to the Android operating system. Mobile phones nowadays evolve as data repositories in which pieces of data belong to different owners and can or must be protected by different security policies. These pieces of data are used on an open environment controlled by a non-specialist user. The dynamic monitoring of information flows is well adapted for protecting information on an embedded system as a mobile phone. Nevertheless the main difficulty relies on the definition of the information flow policy. We proposed a way to define such a policy for the Android operating system.

6.1.2. Detecting Attacks against Data in Web Applications

In [41] we present RRABIDS (Ruby on Rails Anomaly Based Intrusion Detection System) an application level intrusion detection system for applications implemented with the Ruby on Rails framework. This IDS has been developed in the context of a collaborative project funded by ANR and called DALI.

This work aims at detecting attacks against data in the context of web applications. This anomaly based IDS focuses on the modeling of the application profile in the absence of attacks (called normal profile) using invariants. These invariants are discovered during a learning phase. Then, they are used to instrument the web application at source code level, so that a deviation from the normal profile can be detected at run-time. We showed on simple examples how the approach detects well known categories of web attacks that involve a state violation of the application, such as SQL injections. An assessment phase was performed to evaluate the accuracy of the detection provided by the proposed approach. We learned two lessons during this assessment. First this approach provides excellent results in term of false negatives. Second it demonstrates the importance of the learning phase in terms of false positives.

6.1.3. Visualization of Security Events

After having performed in the begining of the year an extensive state of the art of the current visualisation tools dedicated to security, it now clearly appears that there is an important lack of proposals in the context of security data analytics: most of the current visualization proposals build representations for real-time monitoring and only a few of them really allow the user to crawl its data sources in details. Due to this fact, we decided to focus on visualization for security data analytics.

We also built a new visualisation platform in order to lead experiments. Our new directions and the platform have been presented in [20].

6.1.4. Intrusion Detection System Assessement

In [32], we present Netzob [1], a tool dedicated to semi-automatic network protocol reverse-engineering. Such a tool is useful to understand proprietary or non-documented protocols, which is often the case in security analysis or security product assessments. Netzob leverages different algorithms from the fields of bio-informatics and automata theory to infer both the vocabulary and the grammar of undocumented protocols. The vocabulary is inferred from message sequences previously captured (network packets, function call traces, etc.) whereas the grammar inference needs a working implementation of the protocol, which is executed in a confined environment and is used as an oracle. The inferred model could be used to automatically build a client or server implementation of the protocol to generate realistic network traffic.

6.2. Privacy

6.2.1. Geoprivacy

Recent advances in geolocated capacities, secure and verified positioning techniques, ubiquitous connectivity, as well as mobile and embedded systems, have led to the development of a plethora of Location-Based Services (LBS), personalizing the services they deliver according to the location of the user querying the service. However, beyond the benefits they provide, users have started to be worried about the privacy breaches caused by such systems. Among all the Personally Identifiable Information (PII), learning the location of an individual is one of the greatest threats against privacy. In particular, an inference attack [19], can use mobility data (together with some auxiliary information) to deduce the points of interests characterizing his mobility, to predict his past, current and future locations [34] or even to identify his social network.

In order to address and mitigate these privacy issues, within the AMORES project [31], we aim at developing an architecture for the provision of privacy-preserving and resilient collaborative services for "mobiquitous" (*i.e.*, mobile and ubiquitous) systems. The project is built around three uses-cases from the area of publication transportation: (1) dynamic carpooling, (2) real-time computation of multimodal transportation itineraries and (3) mobile social networking. Recently, we have introduced the concept of locanym [35], which corresponds to a pseudonym linked to a particular location that could be used as a basis for developing privacy-preserving LBS.

6.2.2. Privacy-enhanced Social Networks

In [49], we have introduced a new research track focusing on the protection of privacy in distributed social networks, which corresponds to the PhD thesis of Regina Paiva Melo Marin. Our first step has been a study of the needs and practices regarding privacy and personal data policies in social networking frameworks. The commonly accepted requirements for general privacy policies are evaluated with respect to the corresponding notions found in European regulations, and then interpreted in the context of social networking applications. One of the main finding of this study is that some of these requirements are not met by the existing social networks (be they widely used or in development, centralized or distributed, focusing on personal data monetization or on user privacy). The concept of *purpose*, as well as the associated notions of minimization, finality and proportionality, in particular, appears to be insufficiently described in the various policy models. Finally, we have proposed a set of minimal requirements that a privacy policy framework designed for distributed social networks should meet for it to be sufficiently expressive with regards to the current regulations.

[1] http://www.netzob.org

6.2.3. Privacy Enhancing Technologies

Even though they integrate some blind submission functionalities, current conference review systems, such as EasyChair and EDAS, do not fully protect the privacy of authors and reviewers, in particular from the eyes of the program chair. As a consequence, their use may cause a lack of objectivity in the decision process. To address this issue, we have proposed in collaboration with researchers from the Université de Montréal, P3ERS (for Privacy-Preserving PEer Review System) [17], a distributed conference review system based on group signatures, which aims at preserving the privacy of all participants involved in the peer review process. One of the main ideas of P3ERS is to ensure the privacy of both the authors and the reviewers (and this even from the point of view of the conference provider and the conference chair) by using two different groups of users. In particular, the authors can submit anonymized papers on behalf of the author group to the program chair, who then dispatches the papers according to the declared skills of the reviewer group members in an oblivious manner. In this way, the program chair knows neither the identity of the authors (until a paper is accepted, if it is) nor the correspondence between papers and reviewers.

In [25], we have considered the setting in which the profile of a user is represented in a compact way, as a Bloom filter, and the main objective is to privately compute in a distributed manner the similarity between users by relying only on the Bloom filter representation. In particular, our main objective is to provide a high level of privacy with respect to the profile even if a potentially unbounded number of similarity computations take place, thus calling for a non-interactive mechanism. To achieve this, we have proposed a novel non-interactive differentially private mechanism called BLIP (for BLoom-and-flIP) for randomizing Bloom filters. This approach relies on a bit flipping mechanism and offers high privacy guarantees while maintaining a small communication cost. Another advantage of this non-interactive mechanism is that similarity computation can take place even when the user is offline, which is impossible to achieve with interactive mechanisms. Another contribution of this work is the definition of a probabilistic inference attack, called the "Profile Reconstruction attack", that can be used to reconstruct the profile of an individual from his Bloom filter representation. More specifically, we provided an analysis of the protection offered by BLIP against this profile reconstruction attack by deriving an upper and lower bound for the required value of the differential privacy parameter ϵ.

In order to contribute to solve the personalization/privacy paradox, we have proposed a privacy-preserving architecture for one of the state of the art recommendation algorithm, Slope One [36]. More precisely, we designed SlopPy (for *Slope One with Privacy*), a privacy-preserving version of Slope One in which a user never releases directly his personal information (*i.e.*, his ratings). Rather, each user first perturbs locally his information by applying a Randomized Response Technique before sending this perturbed data to a semi-trusted entity responsible for storing it. While there is a trade-off to set between the desired privacy level and the utility of the resulting recommendation, our preliminary experiments clearly demonstrate that SlopPy is able to provide a high level of privacy at the cost of a small decrease of utility.

A privacy-preserving identity card is a personal device device that allows its owner to prove some binary statements about himself (such as his right of access to some resources or a property linked to his identity) while minimizing personal information leakage. As a follow-up of previous works, we have discussed a taxonomy of threats against the card. Finally, we also proposed for security and cryptography experts some novel challenges and research directions raised by the privacy-preserving identity card [50].

6.2.4. Privacy and Data Mining

In [44], [33], we have introduced a novel inference attack that we coined as the reconstruction attack whose objective is to reconstruct a probabilistic version of the original dataset on which a classifier was learnt from the description of this classifier and possibly some auxiliary information. In a nutshell, the reconstruction attack exploits the structure of the classifier in order to derive a probabilistic version of dataset on which this model has been trained. Moreover, we proposed a general framework that can be used to assess the success of a reconstruction attack in terms of a novel distance between the reconstructed and original datasets. In case of multiple releases of classifiers, we also gave a strategy that can be used to merge the different reconstructed datasets into a single coherent one that is closer to the original dataset than any of the simple reconstructed datasets. Finally, we gave an instantiation of this reconstruction attack on a decision tree classifier that was learnt using the algorithm C4.5 and evaluated experimentally its efficiency. The results of this experimentation demonstrate that the proposed attack is able to reconstruct a significant part of the original dataset, thus highlighting the need to develop new learning algorithms whose output is specifically tailored to mitigate the success of this type of attack.

6.2.5. Privacy and Web Services

We have proposed [18] a new model of security policy based for a first part on our previous works in information flow policy and for a second part on a model of Myers and Liskov. This new model of information flow serves web services security and allows a user to precisely define where its own sensitive pieces of data are allowed to flow through the definition of an information flow policy. A novel feature of such policy is that they can be dynamically updated, which is fundamental in the context of web services that allow the dynamic discovery of services. We have also presented an implementation of this model in a web services orchestration in BPEL (Business Process Execution Language) [18].

6.3. Trust

6.3.1. Privacy Preserving Digital Reputation Mechanism

Digital reputation mechanisms have recently emerged as a promising approach to cope with the specificities of large scale and dynamic systems. Similarly to real world reputation, a digital reputation mechanism expresses a collective opinion about a target user based on aggregated feedback about his past behavior. The resulting reputation score is usually a mathematical object, *e.g.* a number or a percentage. It is used to help entities in deciding whether an interaction with a target user should be considered. Digital reputation mechanisms are thus a powerful tool to incite users to trustworthily behave. Indeed, a user who behaves correctly improves his reputation score, encouraging more users to interact with him. In contrast, misbehaving users have lower reputation scores, which makes it harder for them to interact with other users. To be useful, a reputation mechanism must itself be accurate against adversarial behaviors. Indeed, a user may attack the mechanism to increase his own reputation score or to reduce the reputation of a competitor. A user may also free-ride the mechanism and estimate the reputation of other users without providing his own feedback. From what has been said, it should be clear that reputation is beneficial in order to reduce the potential risk of communicating with almost or completely unknown entities. Unfortunately, the user privacy may easily be jeopardized by reputation mechanisms which is clearly a strong argument to compromise the use of such a mechanism. Indeed, by collecting and aggregating user feedback, or by simply interacting with someone, reputation systems can be easily manipulated in order to deduce user profiles. Thus preserving user privacy while computing robust reputation is a real and important issue that we address in ou work [48], [52]. Our proposition combines techniques and algorithms coming from both distributed systems and privacy research domains. Specifically, we propose to self-organize agents over a logical structured graph, and to exploit properties of these graphs to anonymously store interactions feedback. By relying on robust reputation scores functions we tolerate ballot stuffing, bad mouthing and repudiation attacks. Finally, we guarantee error bounds on the reputation estimation score.

6.4. Other Topics Related to Security and Distributed Computing

6.4.1. Network Monitoring and Fault Detection

Monitoring a system is the ability of collecting and analyzing relevant information provided by the monitored devices so as to be continuously aware of the system state. However, the ever growing complexity and scale of systems makes both real time monitoring and fault detection a quite tedious task. Thus the usually adopted option is to focus solely on a subset of information states, so as to provide coarse-grained indicators. As a consequence, detecting isolated failures or anomalies is a quite challenging issue. We propose in [29] to address this issue by pushing the monitoring task at the edge of the network. We present a peer-to-peer based architecture, which enables nodes to adaptively and efficiently self-organize according to their "health" indicators. By exploiting both temporal and spatial correlations that exist between a device and its vicinity, our approach guarantees that only isolated anomalies (an anomaly is isolated if it impacts solely a monitored device) are reported on the fly to the network operator. We show that the end-to-end detection process, *i.e.*, from the local detection to the management operator reporting, requires a logarithmic number of messages in the size of the network.

6.4.2. Metrics Estimation on Very Large Data Streams

In [27] and [28], we consider the setting of large scale distributed systems, in which each node needs to quickly process a huge amount of data received in the form of a stream that may have been tampered with by an adversary. In this situation, a fundamental problem is how to detect and quantify the amount of work performed by the adversary. To address this issue, we propose AnKLe (for Attack-tolerant eNhanced Kullback-Leibler divergence Estimator), a novel algorithm for estimating the KL divergence of an observed stream compared to the expected one. AnKLe com- bines sampling techniques and information-theoretic methods. It is very efficient, both in terms of space and time complexities, and requires only a single pass over the data stream. Experimental results show that the estimation provided by AnKLe remains accurate even for different adversarial settings for which the quality of other methods dramatically decreases. In [26], considering n as the number of distinct data items in a stream, we show that AnKLe is an (ε, δ)-approximation algorithm with a space complexity $\widetilde{\mathcal{O}}(\frac{1}{\varepsilon} + \frac{1}{\varepsilon^2})$ bits in "most" cases, and $\widetilde{\mathcal{O}}(\frac{1}{\varepsilon} + \frac{n-\varepsilon^{-1}}{\varepsilon^2})$ otherwise. To the best of our knowledge, an approximation algorithm for estimating the Kullback-Leibler divergence has never been analyzed before. We go a step further by considering in [51] the problem of estimating the distance between any two large data streams in small-space constraint. This problem is of utmost importance in data intensive monitoring applications where input streams are generated rapidly. These streams need to be processed on the fly and accurately to quickly determine any deviance from nominal behavior. We present a new metric, the *Sketch ☆-metric*, which allows to define a distance between updatable summaries (or sketches) of large data streams. An important feature of the *Sketch ☆-metric* is that, given a measure on the entire initial data streams, the *Sketch ☆-metric* preserves the axioms of the latter measure on the sketch (such as the non-negativity, the identity, the symmetry, the triangle inequality but also specific properties of the f-divergence or the Bregman one). Extensive experiments conducted on both synthetic traces and real data sets allow us to validate the robustness and accuracy of the *Sketch ☆-metric*.

6.4.3. Robustness Analysis of Large Scale Distributed Systems

In [14] we present an in-depth study of the dynamicity and robustness properties of large-scale distributed systems, and in particular of peer-to-peer systems. When designing such systems, two major issues need to be faced. First, population of these systems evolves continuously (nodes can join and leave the system as often as they wish without any central authority in charge of their control), and second, these systems being open, one needs to defend against the presence of malicious nodes that try to subvert the system. Given robust operations and adversarial strategies, we propose an analytical model of the local behavior of clusters, based on Markov chains. This local model provides an evaluation of the impact of malicious behaviors on the correctness of the system. Moreover, this local model is used to evaluate analytically the performance of the global system, allowing to characterize the global behavior of the system with respect to its dynamics and to the presence of malicious nodes and then to validate our approach. We complete this work by considering in [13], the behavior of a stochastic system composed of several identically distributed, but non independent, discrete-time absorbing Markov chains competing at each instant for a transition. The competition consists in determining at each instant, using a given probability distribution, the only Markov chain allowed to make a transition. We analyze the first time at which one of the Markov chains reaches its absorbing state. When the number of Markov chains goes to infinity, we analyze the asymptotic behavior of the system for an arbitrary probability mass function governing the competition. We give conditions for the existence of the asymptotic distribution and we show how these results apply to cluster-based distributed systems when the competition between the Markov chains is handled by using a geometric distribution.

6.4.4. Secure Multiparty Computation in Dynamic Networks

In [37] in collaboration with researchers from EPFL, we consider the problem of securely conducting a poll in synchronous dynamic networks equipped with a Public Key Infrastructure (PKI). Whereas previous distributed solutions had a communication cost of $O(n^2)$ in an n nodes system, we present SPP (Secure and Private Polling), the first distributed polling protocol requiring only a communication complexity of $O(nlog^3 n)$, which we prove is near-optimal. Our protocol ensures perfect security against a computationally-bounded adversary, tolerates $(1/2?\epsilon)n$ Byzantine nodes for any constant $1/2 > \epsilon > 0$ (not depending on n), and outputs the exact value of the poll with high probability. SPP is composed of two sub-protocols, which we believe to be interesting on their own: SPP-Overlay maintains a structured overlay when nodes leave or join the network, and SPP-Computation conducts the actual poll. We validate the practicality of our approach through experimental evaluations and describe briefly two possible applications of SPP: (1) an optimal Byzantine Agreement protocol whose communication complexity is $\Theta(nlogn)$ and (2) a protocol solving an open question of King and Saia in the context of aggregation functions, namely on the feasibility of performing multiparty secure aggregations with a communication complexity of $o(n^2)$.

6.4.5. Agreement Problems in Unreliable Systems

In distributed systems, replication techniques are used to mask occurrences of accidental and malicious failures. To coordinate efficiently the different replicas, different approaches can be adopted (state machine mechanisms, group communication services, ...). Most solutions are based on agreement protocols. The Consensus service has been recognized as a fundamental building block for fault-tolerant distributed systems. Many different protocols to implement such a service have been proposed, however, little effort has been placed in evaluating their performance. We have proposed a protocol designed to solve several consecutive consensus instances in an asynchronous distributed system prone to crash failures and message omissions. The protocol follows the Paxos approach and integrates two different optimizations to reduce the latency of learning a decision value. As one optimization is risky, dynamics triggering criterion are defined to check at runtime if the context seems to be favorable or not. The proposed protocol is adaptive as it tries to obtain the best performance gain depending on the current context. Moreover, it guarantees the persistence of all decision values. Our experimentation results [39] focus on the impact of the prediction of collisions (i.e. , the cases where the use of the risky optimization is counterproductive).

We consider also the problem of approximate consensus in mobile ad hoc networks in the presence of Byzantine nodes. Each node begins to participate by providing a real number called its initial value. Eventually all correct nodes must obtain final values that are different from each other within a maximum value denoted ϵ (convergence property) and must be in the range of initial values proposed by the correct nodes (validity property). Due to nodes' mobility, the topology is dynamic and unpredictable. In [40], [53], we propose an approximate Byzantine consensus protocol which is based on the linear iteration method. Each node repeatedly executes rounds. During a round, a node moves to a new location, broadcasts its current value, gathers values from its neighbors, and possibly updates its value. In our protocol, nodes are allowed to collect information during several consecutive rounds: thus moving gives them the opportunity to gather progressively enough values. An integer parameter Rc is used to define the maximal number of rounds during which values can be gathered and stored while waiting to be used. A novel sufficient and necessary condition guarantees the final convergence of the consensus protocol. At each stage of the computation, a single correct node is concerned by the requirement expressed by this new condition (the condition is not universal as it is the case in all previous related works). Moreover the condition considers both the topology and the values proposed by correct nodes. If less than one third of the nodes are faulty, the condition can be satisfied. We are working on mobility scenarios (random trajectories, predefined trajectories, meeting points) to assert that the condition can be satisfied for reasonable values of Rc.

DANTE Team

6. New Results

6.1. Use of wireless sensor network for Assessing Interactions between Healthcare Workers and Patients under Airborne Precautions

Direct observation has been widely used to assess interactions between healthcare workers (HCWs) and patients but is time-consuming and feasible only over short periods. We used a Radio Frequency Identification Device (RFID) system to automatically measure HCW-patient interactions [14]. The RFID was well accepted by HCWs. This original technique holds promise for accurately and continuously measuring interactions between HCWs and patients, as a less resource-consuming substitute for direct observation. The results could be used to model the transmission of significant pathogens. HCW perceptions of interactions with patients accurately reflected reality.

6.2. Psychological Aspects of Social Communities

Social Network Analysis has often focused on the structure of the network without taking into account the characteristics of the individual involved. In this work [28], [8], we aim at identifying how individual differences in psychological traits affect the community structure of social networks. Instead of choosing to study only either structural or psychological properties of an individual, our aim is to exhibit in which way the psychological attributes of interacting individuals impacts the social network topology. Using psychological data from the myPersonality application and social data from Facebook, we confront the personality traits of the subjects to metrics obtained after applying the C3 community detection algorithm [41] to the social neighborhood of the subjects. We observe that introverts tend to have less communities and hide into large communities, whereas extroverts tend to act as bridges between more communities, which are on average smaller and of varying cohesion.

6.3. Community detection: dynamic, overlapping, fuzzy

Community, a notion transversal to all areas of Social Network Analysis, has drawn tremendous amount of attention across the sciences in the past decades. Numerous attempts to characterize both the sociological embodiment of the concept as well as its observable structural manifestation in the social network have to this date only converged in spirit. No formal consensus has been reached on the quantiffable aspects of community, despite it being deeply linked to topological and dynamic aspects of the underlying social network.

The DANTE team proceeded results on several aspects of community detection is large scale networks.

- Presenting a fresh approach to the evaluation of communities, we introduces and builds upon the cohesion [8], a novel metric which captures the intrinsic quality, as a community, of a set of nodes in a network. The cohesion, defined in terms of social triads, was found to be highly correlated to the subjective perception of communitiness through the use of a large-scale online experiment in which users were able to compute and rate the quality of their social groups on Facebook. The use of the cohesion proves invaluable in that it offers non- trivial insights on the network structure and its relation to the associated semantic. The use of the cohesion was use for example in order to study Agreement Groups in the United States Senate [35].
- Overlapping community detection is a popular topic in complex networks. As compared to disjoint community structure, overlapping community structure is more suitable to describe networks at a macroscopic level. Overlaps shared by communities play an important role in combining different communities. In this paper, two methods are proposed to detect overlapping community structure. One is called clique optimization, and the other is named fuzzy detection. Clique optimization aims at detecting granular overlaps. The clique optimization method is a fine grain scale approach. Each

granular overlap is a node connected to distinct communities and it is highly connected to each community. Fuzzy detection is at a coarser grain scale and aims at identifying modular overlaps. Modular overlaps represent groups of nodes that have high community membership degrees with several communities. A modular overlap is itself a possible cluster/sub-community [7], [38].

6.4. Structure of Changes in Dynamic Contact Networks

We present a methodology to investigate the structure of dynamic networks in terms of concentration of changes in the network. We handle dynamic networks as series of graphs on a set of nodes and consider the changes occurring between two consecutive graphs in the series. We apply our methodology to various dynamic contact networks coming from different contexts and we show that changes in these networks exhibit a non-trivial structure: they are not spread all over the network but are instead concentrated around a small fraction of nodes. We compare our observations on real-world networks to three classical dynamic network models and show that they do not capture this key property [31].

6.5. Dynamic Resource Management in Clouds: A Probabilistic Approach

Dynamic resource management has become an active area of research in the Cloud Computing paradigm. Cost of resources varies significantly depending on configuration for using them. Hence efficient management of resources is of prime interest to both Cloud Providers and Cloud Users. In this work we suggest a probabilistic resource provisioning approach that can be exploited as the input of a dynamic resource management scheme. Using a Video on Demand use case to justify our claims, we propose an analytical model inspired from standard models developed for epidemiology spreading, to represent sudden and intense workload variations. We show that the resulting model verifies a Large Deviation Principle that statistically characterises extreme rare events, such as the ones produced by ?buzz/flash crowd effects? that may cause workload overflow in the VoD context. This analysis provides valuable insight on expectable abnormal behaviours of systems. We exploit the information obtained using the Large Deviation Principle for the proposed Video on Demand use-case for defining policies (Service Level Agreements). We believe these policies for elastic resource provisioning and usage may be of some interest to all stakeholders in the emerging context of cloud networking [4], [24].

6.6. Classification of Content and Users in BitTorrent by Semi-supervised Learning Methods

P2P downloads still represent a large portion of today?s Internet traffic. More than 100 million users operate BitTorrent and generate more than 30% of the total Internet traffic. Recently, a significant research effort has been done to develop tools for automatic classification of Internet traffic by application. The purpose of the present work is to provide a framework for sub-classification of P2P traffic generated by the BitTorrent protocol. The general intuition is that the users with similar interests download similar contents. This intuition can be rigorously formalised with the help of graph based semi- supervised learning approach. We have chosen to work with PageRank based semi-supervised learning method, which scales well with very large volumes of data. We provide recommendations for the choice of parameters in the PageRank based semi-supervised learning method. In particular, we show that it is advantageous to choose labelled points with large PageRank score.

This work was awarded best paper at the 3rd International Workshop on Traffic Analysis and Classification (in conjunction with the 8th International Wireless Communications and Mobile Computing Conference, 2012) [21] and led to a companion paper [22].

6.7. Large deviations estimates for the multiscale analysis of heart rate variability

In the realm of multiscale signal analysis, multifractal analysis provides with a natural and rich framework to measure the roughness of a time series. As such, it has drawn special attention of both mathematicians and practitioners, and led them to characterize relevant physiological factors impacting the heart rate variability. Notwithstanding these considerable progresses, multi- fractal analysis almost exclusively developed around the concept of Legendre singularity spectrum, for which efficient and elaborate estimators exist, but which are structurally blind to subtle features like non-concavity or, to a certain extent, non scaling of the distributions. Large deviations theory al- lows bypassing these limitations but it is only very recently that performing estimators were proposed to reliably compute the corresponding large devia- tions singularity spectrum. In this article, we illustrate the relevance of this approach, on both theoretical objects and on human heart rate signals from the Physionet public database. As conjectured, we verify that large devia- tions principles reveal significant information that otherwise remains hidden with classical approaches, and which can be reminiscent of some physiolog- ical characteristics. In particular we quantify the presence/absence of scale invariance of RR signals.

These results gather most achievements we carried out within the ANR project DMASC.

6.8. An Inexpensive Packet Capture Solution with Robust and Accurate Timestamping

The availability of inexpensive and reliable packet capture solutions is highly desirable for the management of future Internet infrastructures and practices. Currently, available solutions are either 1) based on GPS antennas and dedicated hardware and thus are expensive and difficult to deploy, or 2) based on commodity hardware and standard synchronization protocols and thus have inaccurate timestamps and cannot handle monitoring at high rate. In a series of ongoing works in collaboration with the Melbourne University (Australia), we proposed an architecture for a packet monitoring solution which combines inexpensive network cards capable of hardware timestamping, with RAD- clock, an open source software clock. In different papers, we presented the first implementation and evaluation of our approach, demonstrating a good compromise between affordability and accuracy [33], [36].

6.9. KBAC: Knowledge-Based Admission Control

Many methods have been proposed in the literature to perform admission control in order to provide a sufficient level of Quality of Service (QoS) to accepted flows. In this work, we introduce a novel data-driven method based on a timevarying model that we refer to as Knowledge-Based Admission Control solution (KBAC). Our KBAC solution consists of three main stages: (i) collect measurements on the on-going traffic over the communication link; (ii) maintain an up-to-date broad view of the link behavior, and feed it to a Knowledge Plane; (iii) model the observed link behavior by a mono-server queue whose parameters are set automatically and which predicts the expected QoS if a flow requesting admission were to be accepted. Our KBAC solution provides a probabilistic guarantee whose admission threshold is either expressed, as a bounded delay or as a bounded loss rate. We run extensive simulations to assess the behavior of our KBAC solution in the case of a delay threshold. The results show that our KBAC solution leads to a good trade-off between flow performance and resource utilization. This ability stems from the quick and automatic adjustment of its admission policy according to the actual variations on the traffic conditions [19].

6.10. Substitution Networks: Performance Collapse due to Overhead in Communication Times

A substitution network is a wireless solution whose purpose is to bring back connectivity or to provide additional bandwidth capacity to a network that just suffered a failure or a dramatic surge in its workload. We analyze the performance of the simplest possible multihop topology for a substitution network, i.e., the multihop chain subject to traffic transmitted in both directions. Clearly, the potential capacity of a substitution network, whose technology should be embedded in mobile routers, is very likely to be far much smaller than the prior base network. We investigate the actual performance attained by such a substitution network under various conditions of the chain length and the carrier sensing range. Our results show that the capacity, viz. its maximum attainable throughput, reaches a peak at a given workload and then, for larger values of workload, decreases towards an asymptote which value can be drastically lower than the peak value. We give insights into this performance collapse and show the need for a suitable admission control [18].

6.11. Characterisation and Application of Idle Period Durations in IEEE 802.11 DCF-based Multihop Wireless Networks

Multihop wireless networks are used to provide internet connectivity to the users and the level of performance and quality expected by these users are increasing. In order to meet these performance and quality requirements, wireless communications should be enhanced. Previous works from the literature show that the performance and quality provided by an IEEE 802.11-based multihop wireless network are far from optimal and that there exist different ways to increase the efficiency and the quality of service of such a network. Some studies show that using the medium state as a parameter to tune the behaviour of an IEEE 802.11-based multihop network is an appropriate way to proceed. A station in a IEEE 802.11-based multihop wireless network senses the medium either busy or idle. The durations of idle periods and busy periods and their distributions have a clear impact on the network and nodes performance. The understanding of the relationship between these indicators, namely idle and busy periods, the network topology and the traffic, would give new insights to enhance the performance and quality of multihop wireless networks. Due to its multihop and distributed nature, the characterisation of idle period durations is difficult in such a network. This work explores the characterisation of idle period distribution by proposing a new analytical model and provides an application of this characterisation with the design of an adaptive backoff algorithm based on idle periods [30].

DIONYSOS Project-Team

5. New Results

5.1. Quality of Experience

Participants: Gerardo Rubino, Adlen Ksentini, Yassine Hadjadj-Aoul, Sofiene Jelassi, Sebastián Basterrech.

We continue the development of the PSQA technology (Pseudo-Subjective Quality Assessment) in the area of Quality of Experience (QoE). PSQA is today a stable technology allowing to build measuring modules capable of quantifying the quality of a video or an audio sequence, as perceived by the user, when received through an IP network. It provides an accurate and efficiently computed evaluation of quality. Accuracy means that PSQA gives values close to those than can be obtained from a panel of human observers, under a controlled subjective testing experiment, following an appropriate standard (which depends on the type of sequence or application). Efficiency means that our measuring tool can work in real time, if necessary. Observe that perceived quality is the main component of QoE. PSQA works by analyzing the networking environment of the communication and some the technical characteristics of the latter. It works without any need to the original sequence (as such, it belongs to the family of *no-reference* techniques).

It must be pointed out that a PSQA measuring or monitoring module is network dependent and application dependent. Basically, for each specific networking technology, application, service, the module must be built from scratch. But once built, it works automatically and efficiently, allowing if necessary its use in real time.

At the heart of the PSQA approach there is the statistical learning process necessary to develop measuring modules. So far we have been using Random Neural Networks (RNNs) as our learning tool (see [96] for a general description), but recently, we have started to explore other approaches. For instance, in the last ten years a new computational paradigm was presented under the name of *Reservoir Computing* (RC) [93] covering the main limitations in training time for recurrent neural networks while introducing no significant disadvantages. Two RC models have been developed independently and simultaneously under the name of *Liquid State Machine* (LSM) [95] and *Echo State Networks* (ESN) [93] and constitute today one of the basic paradigms for Recurrent Neural Networks modeling [94]. The main characteristic of the RC model is that it separates two parts: a static sub-structure called *reservoir* which involves the use of cycles in order to provide dynamic memory in the network, and a parametric part composed of a function such as a multiple linear regression or a classical single layer network. The reservoir can be seen as a dynamical system that expand the input stream in a space of states. The learning part of the model is the parametric one. In [38] we propose a new learning tool which merges the capabilities of Random Neural Networks (RNNs) with those of Reservoir Computing Models (RCMs). We keep some of the nice features of RNNs with the ability of RCMs in predicting time series values. Our tool is called Echo State Queueing Network. In the paper, we illustrate its performances in predicting, in particular, Internet traffic. We also worked on the bottleneck of the PSQA building process, from the time consuming point of view, the subjective test sessions. We proposed in [49] and [48] new PSQA modules for VoIP and SVC video, respectively. In [49], we used PESQ for replacing the subjective test in the training step of PSQA. This module is dedicated to iLBC and Speex codecs. Whereas in [48], we used VQM tool to evaluate the SVC video sequences to train PSQA.

In [31], a general presentation of our approach in Dionysos was given, together with some guidelines in looking for extensions able to deal with the evaluation of generic applications or services over the Internet.

We presented a tutorial on Quality of Experience in Qest'2012 [69], based on our past research results in evaluating the perceptual quality in voice or video applications, and on the current work performed in the QuEEN project.

Our perceptual quality work is being extended to investigate the quality of user experience including a large scope that involves human and technology factors. This work is conducted in the context of the Celtic-QuEEN project where a complete QoE monitoring platform is being designed. In Qest'2012 [69], we presented a tutorial on Quality of Experience based on our past research results in evaluating the perceptual quality in voice or video applications, and on the current work performed in QuEEN.

On the other hand, we continue our study of quality of temporally interrupted VoIP service frequently observed over wireless and data networks. A flagship paper regarding the perception of interruptions in the context of VoIP service is published in [53]. In [21] we presented a detailed state-or-the-art in the area.

5.2. Network Economics

Participants: Bruno Tuffin, Jean-Marc Vigne.

While pricing telecommunication networks was one of our main activities for the past few years, we are now dealing with the more general topic of *network economics* (see for instance [83]). We have tackled it from different sides: i) investigating how QoS or QoE can be related to users' willingness to pay, ii) investigating the consequences and equilibria due competition among providers in different contexts, iii) looking at the economics of applications, for example adword auctions for search engines, iv) studying the network neutrality issue, and v) the not so considered problem of search-neutrality.

On the first item, we have studied in [78] how utility functions can be related to QoE recent research. Indeed, a logarithmic version of utility usually serves as the standard example due to its simplicity and mathematical tractability. We argue that there are much more (and better) reasons to consider logarithmic utilities as really paradigmatic, at least when it comes to characterizing user experience with specific telecommunication services. We justify this claim and demonstrate that, especially for Voice-over-IP and mobile broadband scenarios, there is increasing evidence that user experience and satisfaction follows logarithmic laws. Finally, we go even one step further and put these results into the broader context of the Weber-Fechner Law, a key principle in psychophysics describing the general relationship between the magnitude of a physical stimulus and its perceived intensity within the human sensory system.

A notable part of our activity has been related to competition among telecommunication providers, mainly within the framework of the ANR CAPTURES project ending this year. The goal is to improve most of the pricing models analysis which only deal with a single provider while competition (that is observed in the telecommunication industry) can drive to totally different outcomes. A general view of some of our results is summarized in [77]. A general model of competition in loss networks is described and analyzed in [25] as a two-levels game: at the smallest time scale, users' demand is split among providers according to Wardrop principle, depending on the access price and available QoS (depending itself on the level of demand at the provider); at the largest time scale, providers play a pricing game, trying non-cooperatively to maximize their revenue. A striking result is that this game leads to the same outcome than if providers were cooperatively trying to maximize social welfare: the so-called *price of anarchy* is equal to one. In [59], we present a similar model of competition on prices between two telecommunication service providers sharing an access resource, which can for example be a single WiFi spectrum. We again obtain a two-level game corresponding to two time scales of decisions: at the smallest time scale, users play an association game by choosing their provider (or none) depending on price, provider reputation and congestion level; at the largest time scale, providers compete on prices. We show that the association game always has an equilibrium, but that several equilibria can exist. The pricing game is then solved by assuming that providers are risk-averse and try to maximize the minimal revenue they can get at a user equilibrium. We illustrate what can be the outcome of this game and that there are situations for which providers can co-exist.

Network economics is not only about ISPs, it also deals with the application side. In order to make money, many service providers base their revenue on advertisement. Search engines for example get revenue thanks to adword auctions, where commercial links are proposed and charged to advertisers as soon as the link is clicked through. The strategies of the search engine and advertisers are described and analyzed in [24].

A new issue on which most of our work has focused in 2012 is related to the *network neutrality debate*. This debates comes from the increasing traffic asymmetry between Internet Service Providers (ISPs), mainly due to some prominent and resource consuming content providers (Cps) which are usually connected to a single ISP. Thus the ISPs to whom those CPs are not directly connected have started to wonder why distant CPs should not be charged by them, with the threat of their traffic not being delivered if they do not accept to pay, or their quality of service decreased. In [79], we have described and analyzed the respective arguments of neutrality proponents and opponents, and we have also participated to Inria's response to the ARCEP consultation on the topic [90]. We have reviewed in [50], [85] the economic transit agreements between ISPs in order to determine their best strategy. We have defined a model with two ISPs, each providing direct connectivity to a fixed proportion of the content and competing in terms of price for end users, who select their ISP based on the price per unit of available content. We have analyzed and compared, thanks to game-theoretic tools, three different situations: the case of peering between the ISPs, the case where ISPs do not share their traffic (exclusivity arrangements), and the case where they fix a transfer price per unit of volume. The impact on the network neutrality debate is then discussed. An analysis with a hierarchy of providers, with separated backbone providers and access providers, is performed in [89]. We also remarked that while there have been many studies discussing the advantages and drawbacks of neutrality, there is no game-theoretical work dealing with the observable situation of competitive ISPs in front of a (quasi-)monopolistic CP. Though, this is a typical situation that is condemned by ISPs and, according to them, another reason of the non-neutrality need. We have developed and analyzed in [40], [84] two different models describing the relations between two competitive ISPs and a single CP, played as a three-level game corresponding to three different time scales. At the largest time scale, side payments (if any) are determined. At a smaller time scale, ISPs decide their (flat-rate) subscription fee (toward users), then the CP chooses the (flat-rate) price to charge users. Users finally select their ISP (if any) using a price-based discrete choice model in [84] or following Wardrop principle in [40], and decide whether to also subscribe to the CP service. The game is analyzed by backward induction. As a conclusion, we obtain among other things that non-neutrality may be beneficial to the CP, and not necessarily to ISPs, unless the side payments are decided by ISPs (through a non-cooperative game). Another specific scenario is studied in [51], where the impact of wholesale prices is examined in a context where the end customer access both free content and pay-per-use content, delivered by two different providers through a common network provider. We formulate and solve the game between the network provider and the pay-per-use content provider, where both use the price they separately charge the end customer with as a leverage to maximize their profits. In the neutral case (the network provider charges equal wholesale prices to the two content providers), the benefits coming from wholesale price reductions are largely retained by the pay-per-use content provider. When the free content provider is charged more than its pay-per-use competitor, both the network provider and the pay-per-use content provider see their profit increase, while the end customer experiences a negligible reduction in the retail price.

If network neutrality has recently attracted a lot of attention, *search neutrality* is also becoming a vivid subject of discussion because a non-neutral search may prevent some relevant content from being accessed by users. We propose in [88] to model two situations of a non-neutral search engine behavior, which can rank the link propositions according to the profit a search can generate for it, instead of just relevance: the case when the search engine owns some content, and the case when it imposes a tax on organic links, a bit similarly to what it does for commercial links. We analyze the particular (and deterministic) situation of a single keyword, and describe the problem for the whole potential set of keywords. In [52], we analyze one behavior that results in search bias: the payment by content providers to the search engine in order to improve the chances to be located (and accessed) by a search engine user. A simple game theory-based model is presented, where both a search engine and a content provider interact strategically, and the aggregated behavior of users is modeled by a demand function. The output of each stakeholder when the search engine is engaged in such a non-neutral behavior is compared with the neutral case when no such side payment is present.

5.3. Wireless Networks

Participants: Adlen Ksentini, Yassine Hadjadj-Aoul, Bruno Sericola.

Long Term Evolution (LTE) represents the next generation of Cellular networks or 4G. It allows increasing the data rate and hence services that can be proposed to users. A notable part of activity in cellular networks and particularly in LTE, is related to increasing the user QoE. Due to their numerous advantages, current trends show a growing number of femtocell deployments. However, femtocells would become less attractive to the general consumers if they cannot keep up with the service quality that the macro cellular network should provide. Given the fact that the quality of mobile services provided at femtocells depends largely on the level of congestion on the backhaul link, in [71] we introduced a flow mobility/handover admission control method that makes decisions on layer-three handovers from macro network to femtocell network and/or on entire or partial flow mobility between the two networks based on predicted QoS taking into account metrics such as network load/congestion indications and based on predicted QoE metrics. In [70], we proposed a complete framework that anticipates QoS/QoE (Quality of Experience) degradation and proactively defines policies for LTE-connected cars (UEs) to select the most adequate radio access out of WiFi and LTE. For a particular application, the proposed framework considers the application type, the mobility feature (e.g., speed, user mobility entire/partial path, user final/intermediate destination), and the traffic dynamics over the backhauls of both LTE and WiFi networks in order to predict and allow the UE to select the best network that maximizes user QoE throughout the mobility path.

In [33],[23] we considered LTE networks as candidates for hosting the Machine to Machine communication (or Machine Type Communication in the 3GPP jargon). One of the most important problems posed by this kind of traffic is congestion. Congestion concerns all the parts of the network, both the radio and the core networks impacting both the user data and the control planes. In these works, we proposed a congestion aware admission control solution that selectively rejects signaling messages from MTC devices at the radio access network following a probability that is set based on a proportional integrative derivative (PID) controller (from control theory) reflecting the congestion level of a relevant core network node.

Another part of our activities in wireless network are related to energy saving. Indeed, one of the biggest problem today in the wireless world is that wireless devices are battery driven, which reduce their operating lifetime. We addressed the energy issue in wireless network for two different contexts: (i) rich media (such as VoIP) delivery in Wireless LAN; (ii) Wireless Sensor Network (WSN).

In WLAN, mobile stations conserve energy by maximizing the sleep mode periods of the wireless interfaces. Despite of its efficiency, this mode is incompatible with real-time applications and media streaming, like VoIP. In fact, maximizing the sleep mode periods is directly translated into an increased delay, which induces packets losses when exceeding certain thresholds (e.g. buffer overflow and late packet loss), and may severely degrade the perceived user's QoE. We first review a clear state of the art on energy saving for mobiles communication [22]. Then, in [56], we showed the relation between user QoE and the sleep period in the context of Voice over Wireless Lan (VoWLAN). The system was modeled and controlled using a PID controller, which computes the sleep period enabling to reach a QoE reference value. Thus, we achieved the trade-off between energy consumption and QoE.

On the other hand, Wireless Sensor Networks (WSN) protocols focus primarily on power conservation, because of the limited capacity of the sensor nodes' batteries. In [64] we addressed the case of using radio diversity in WSN (more than one antenna). In this work, we proposed a scheme for radio diversity that can balance, depending on the traffic nature in the network, between minimizing the energy consumption or minimizing the end-to-end delay. The proposed scheme combines the benefit of two metrics, which aim separately to minimize the energy consumption, and to minimize delay when delivering packets to the end-user. In [57], we worked on the localization problem in WSN by introducing a new way to determine the sensors' residence area. Our new localization algorithm is based on the geometric shape of half-symmetric lens. In [81] we developed a performance analysis of a compression scheme designed to save energy, for specific types of WSN.

In [55], we presented the DVB-T2 simulation module for OPNET. Note that this module is the only available implementation of DVB-T2 in network simulators.

5.4. Information-Centric Networks

Participants: Yassine Hadjadj-Aoul, Gerardo Rubino, Leila Ghazzai.

The rise of popularity of video streaming services has resulted in increased volumes of network traffic, which in turn have created Internet bottlenecks leading to perceived quality degradations. One of the recognized good ways to tackle this type of congestion is to make the contents available inside ISPs' networks. We thus proposed, in [73] a network-friendly content delivery architecture that considers the complex video distribution chain and its associated business models. This comprehensive architecture allows a network operator to fully engineer video traffic distribution in order to both alleviate peering links' workload and improve delivered QoS. This proposal is fully compatible with Adaptive Bitrate Streaming (ABS) architectures, which are currently used to distribute video in the Internet.

The Content providers are increasingly becoming interested in evaluating the performance of such streaming protocol from the final users' perspective. Indeed, more importance is being attached to the quality as perceived by the final users, or Quality of Experience (QoE), as compared to just Quality of Service. Thus, we addressed in [68] the problem of estimating the QoE of video streaming in TCP/IP networks. As a solution, we designed an automatic no-reference QoE estimation module for HTTP video streaming using TCP and H.264 video codec. The proposed approach is different from the existing ones as it addresses the problem of measuring QoE in the combined case of adaptive video bitrates and the use of a reliable transport protocol. This is the case of the adaptive streaming over HTTP.

On the other hand, as introduced by ICN's content caching mainly addresses the management of the content in a particular cache, while the content replication consists in disseminating data in its way to the destination. The benefits of contents' replication can be completely cancelled with a bad caching technique. Thus, we proposed, in [75], to analyse the interaction existing between caching strategies and content replication.

5.5. Interoperability assessment and Internet of Things

Participants: César Viho, Nanxing Chen, Anthony Baire.

The Internet of Things (IoT) brings new challenges to interoperability assessment by introducing the necessity to deal with non reliable environments connecting plenty billions of objects widely distributed. In this context, the IETF Constrained Application Protocol (CoAP) has been designed, which is an application-layer protocol on keeping in mind the various issues of constrained environment to realize interoperations with constrained networks and nodes.

As one of the most important protocol for the future Internet of Things, the number of smart objects using CoAP is expected to grow substantially. For CoAP applications to be widely adopted by the industry, interoperability testing is required to ensure that CoAP implementations from different vendors work well together. Therefore, in the recent period, we propose an interoperability testing methodology using a *passive* approach. Contrary to the classical testing method used in conventional interoperability testing events, which is done by actively stimulating the implementations and verifying the outputs, we apply passive testing. It is a technique based only on observation [47]. Its non-intrusive nature makes it appropriate for interoperability testing, especially in the context of IoT. We have also developed a tool that implement this passive method that has been used successfully to test CoAP implementations during the two CoAP Plugtest interoperability sessions organized by ETSI and IPSO Alliance [44]. Our contributions and originality of this work published in [46] are three-fold: *(i)* A new testing method using a passive approach. *(ii)* As IoT implies providing services in lossy networks, we also take into account fundamental CoAP implementations interoperability testing in lossy context. *(iii)* Contrary to manual verification used in conventional interoperability testing events, the verification procedure has been automatized by a test validation tool, which increases the test efficiency while reducing testing time and costs.

5.6. Performance Evaluation of Distributed Systems

Participants: Bruno Sericola, Gerardo Rubino, Laura Aspirot, Romaric Ludinard.

In [92] and [13], we consider the behavior of a stochastic system composed of several identically distributed, but non independent, discrete-time absorbing Markov chains competing at each instant for a transition. The competition consists in determining at each instant, using a given probability distribution, the only Markov chain allowed to make a transition. We analyze the first time at which one of the Markov chains reaches its absorbing state. We obtain its distribution and its expectation and we propose an algorithm to compute these quantities. We also exhibit the asymptotic behavior of the system when the number of Markov chains goes to infinity. Actually, this problem comes from the analysis of large-scale distributed systems and we show how our results apply to this domain.

In [14], we present an in-depth study of the dynamicity and robustness properties of large-scale distributed systems, and in particular of peer-to-peer systems. When designing such systems, two major issues need to be faced. First, population of these systems evolves continuously (nodes can join and leave the system as often as they wish without any central authority in charge of their control), and second, these systems being open, one needs to defend against the presence of malicious nodes that try to subvert the system. Given robust operations and adversarial strategies, we propose an analytical model of the local behavior of clusters, based on Markov chains. This local model provides an evaluation of the impact of malicious behaviors on the correctness of the system. Moreover, this local model is used to evaluate analytically the performance of the global system, allowing to characterize its global behavior with respect to its dynamics and to the presence of malicious nodes, and then to validate our approach.

Monitoring a system is the ability of collecting and analyzing relevant information provided by the monitored devices so as to be continuously aware of the system's state. However, the ever growing complexity and scale of systems makes both real time monitoring and fault detection a quite tedious task. The usually adopted option is to focus solely on a subset of information states, so as to provide coarse-grained indicators. As a consequence, detecting isolated failures or anomalies is a quite challenging issue. In [34], we propose to address this issue by pushing the monitoring task at the edge of the network. We present a peer-to-peer-based architecture, which enables nodes to self-organize according to their "health" indicators. By exploiting both temporal and spatial correlations that exist between a device and its vicinity, our approach guarantees that only isolated anomalies (an anomaly is isolated if it impacts solely a monitored device) are reported on the fly to the network operator. We show that the end-to-end detection process, i.e., from the local detection to the management operator reporting, requires a logarithmic number of messages in the size of the network. This work led to the patent [91] with Technicolor.

In [66] we continued previous efforts in the design of peer-to-peer networks for transmitting video content. In the past, we develop tools allowing a perceptual quality-based design tool. In [66], we explore an architectural idea where the video stream is decomposed in sequential sets of chunks that we call "windows". The paper explores some aspects of the performance of such a transmission scheme. The techniques used are Markovian models which are simulated, and deterministic dynamical systems that allow for some equilibrium analysis.

5.7. Monte Carlo

Participants: Bruno Tuffin, Gerardo Rubino, Pablo Sartor.

We maintain a research activity in different areas related to dependability, performability and vulnerability analysis of communication systems, using both the Monte Carlo and the Quasi-Monte Carlo approaches to evaluate the relevant metrics. Monte Carlo (and Quasi-Monte Carlo) methods often represent the only tool able to solve complex problems of these types. However, when the events of interest are rare, simulation requires a special attention, to accelerate the occurrence of the event and get unbiased estimators of the event of interest with a sufficiently small relative variance. This is the main problem in the area. Dionysos' work focuses then in dealing with the rare event situation.

In [72] we have overviewed how the zero-variance importance sampling can be approximated in classical reliability problems. In general, we look for estimators such that the relative accuracy of the output is "controlled" when the rarity is getting more and more critical. Different robustness properties of estimators have been defined in the literature. However, these properties are not adapted to estimators coming from a parametric family for which the optimal parameter is random due to a learning algorithm. These estimators have random accuracy. For this reason, we motivate in [65] the need to define probabilistic robustness properties. We especially focus on the so-called probabilistic bounded relative error property. We additionally provide sufficient conditions, both in general and in Markov settings, to satisfy such a property, and hope that it will foster discussions and new works in the area.

In [43] and [18] we present results concerning the evaluation using Monte Carlo techniques, of a specific reliability metric for communication networks, based not only on connectivity properties, as in the classical network reliability measure, but also in the lengths of the paths. In [43], we propose bounds of the metric that can be used to derive a variance reduction technique. In [18], we describe techniques to analyze what could be called performability aspects of networks also based on the number of hops between sources and terminals. Let us also mention here our publication [16], where we discuss the exact computation of these new types of metrics, and [29], where other related combinatorial problems are discussed (here, optimization problems also based on connectivity properties, from the design point of view). In [17], we propose a new version of the RVR principle, leading to a variance reduction technique for the classic network reliability problem. Paper [28] proposes a splitting algorithm for the same problem. The approach is quite straightforward, after the static problem is transformed into a dynamic one using the well known Creation Process. In [42] we explore a very general conditioning-based approach, including as a particular case the family of splitting procedures. We explore this idea through the analysis of dependability properties of complex systems using Markov models.

When looking specifically at static network reliability models, as described in the previous paragraph, it is often typically assumed that the failures of their components are independent. This assumption allows for the design of efficient Monte Carlo algorithms that can estimate the network reliability in settings where it is a rare-event probability. Despite this computational benefit, independent component failures is frequently not a realistic modeling assumption for real-life networks. In [39] we show how the splitting methods for rare-event simulation can be used to estimate the reliability of a network model that incorporates a realistic dependence structure via the Marshal-Olkin copula.

In [15], we present a versatile Monte Carlo method for estimating multidimensional integrals, with applications to rare-event probability estimation. The method fuses two distinct and popular Monte Carlo simulation methods, Markov chain Monte Carlo and importance sampling, into a single algorithm. We show that for some applied numerical examples the proposed Markov Chain importance sampling algorithm performs better than methods based solely on importance sampling or MCMC.

Finally, in two presentations [67] and [32] we discuss the main problems when analyzing rare events using Monte Carlo methods, focusing on robustness properties of the corresponding estimators.

5.8. Analytic models

Participants: Bruno Sericola, Gerardo Rubino, Raymond Marie, Laura Aspirot.

Fluid models are powerful tools for evaluating the performance of packet telecommunication networks. By masking the complexity of discrete packet based systems, fluid models are in general easier to analyze and yield simple dimensioning formulas. Among fluid queuing systems, those with arrival rates modulated by Markov chains are very efficient to capture the burst structure of packet arrivals, notably in the Internet because of bulk data transfers. By exploiting the Markov property, very efficient numerical algorithms can be designed to estimate performance metrics such the overflow probability, the delay of a fluid particle or the duration of a busy period. In [76], we analyze the transient behaviour of a fluid queue driven by a general ergodic birth and death process using spectral theory in the Laplace transform domain. These results are applied to the stationary regime and to the busy period analysis of that fluid queue.

In [36], another type of fluid model is considered. We present preliminary results on the analysis of a Machine Repairman Model when the number of machines goes to infinity. The analysis is based on identifying appropriate fluid limits of the associated stochastic processes. We are currently working on the analysis of the speed of the convergence of these stochastic processes towards their fluid limits.

In [19], we present an approximate method for the transient analysis of stiff CTMC. The origin of the method is due to S. M. Ross who proposed to approximate the transient probability at a deterministic time t by the value of the transient probability at a random time X where X is an Erlang random variable having expectation t. The major contributions of the paper are the use of new numerical techniques to solve the basic equations of the original method and the extension of the method to reward measures. We also conduct an experimental evaluation of the resulting errors using non-trivial examples.

In [86], we presented an extension of ROBDDs that is able to accommodate certain dependencies among their (Boolean) variables. In particular, this extension shows evidence of being applicable to evaluating the dependability (reliability, availability) of systems whose structures are representable by a Boolean function. This extension consists of three main parts. The first part is the notion of a phratry with its associated new definitions and constraints. The second part consists of the adaptation and complementation of the original rules used in the construction of ROBDDs. The final part concerns additional custom-made steps needed to determine the functional valuations that are specific to solving measure in question.

The survivability of a system being its ability to function during and after a failure, we developed in [63] a model to study the power distribution in smart grids during the (transient) period that starts after a failure till the system fully recovers. The proposed model bridges power flow modeling of reactive power compensation with performability/survivability modeling of automation distribution networks. We use a Markov chain to characterize the phased recovery of the system after a failure. Then, we associate with each state of the Markov chain a set of corresponding rewards to characterize the active and reactive power supplied and demanded in that state. We connect the survivability model with an availability model, to produce a generalization of the System Average Interruption Duration Index (SAIDI) and the Customer Average Interruption Duration Index (CAIDI), which are two of the most important power grid reliability metrics. The survivability model allows us to obtain closed form expressions for the SAIDI and related metrics.

In [62], we consider the case of important systems located on operational sites far away from logistic support forces, either because the operational site is in an inhospitality place, or because it is not profitable to maintain a dedicated team on the operational site. Due to the importance of the systems, some service level agreement has been signed, including conditional financial clauses. To take into account such a situation, a preventive maintenance is realized according to projected calendars. The paper shows that, given that the life-times of equipments are supposed to be Erlang-k distributed, it is optimal to realize a preventive maintenance, as long as the ratio of the two intervention costs C_p/C_c is lower than the ratio $(k-1)/k$, C_p being the cost of a preventive maintenance intervention and C_c being the cost of a curative maintenance intervention (because of excessive delay, there is a significant penalty associated with each curative maintenance intervention). The methodology to compute the optimal value of the period T^* and the corresponding optimal cost per time unit are presented, for a given value of the ratio C_p/C_c. An extended version of this work has been accepted for publication in a journal ([26]).

The study [60] focuses on the determination of the probability distributions of two random variables, the asymptotic "up-time" and "down-time" of a system for the sake of potential "Service Level Agreement". In these new generation agreements, penalizations can be enforced for a too long "down-time" or for a too short "up-time". First, we determine the probability distributions of the two random variables "up-time" and "down-time", for a system with a general structural function. Second, we point out the importance of rare events such as the backorders in the contribution of a large tail distribution of the down-time. Respectively, we exhibit the importance of redundant structures and also of sub-system hyper-exponential lifetimes in the existence of short up-times, with respect to the mean up-time value of the system.

The study [61] deals with the determination of spares of systems of systems of the same type (such as fleet of aircraft, fleet of ship). For a multi-site workshop and multi-level of repair organization, we present an optimization algorithm using the criteria of expected number of backorders as local objective. With respect to a previous algorithm based only on the criteria of the global availability of the system, the new algorithm is, for large maintenance systems, very efficient, in terms of execution time and in of data manipulation.

The study [41] concerns the performance evaluation of crisis management systems with respect to the dimensioning of the system. By definition, a crisis has no steady state and the study must be done on the transient behavior. A faithful model was built (in ALTARICA) and solved thanks to simulation. Our own participation was mainly to determine the number of objects to create such that the simulation ends successfully with a high probability, before running out of available objects.

Last, in [54] we continue the exploration of the concept of duality proposed by Anderson, applied to the analysis of the transient behavior of queueing systems. This work analyzes the transient distribution of the number of customers in a Restart Markovian queue, where together with “typical” customers other signals arrive to the queue having as a consequence the removal of all the customers present in the system.

DISTRIBCOM Project-Team

6. New Results

6.1. Fundamental results and algorithms: distributed planning

Participants: Eric Fabre, Loig Jézéquel.

A planning problem consists in organizing some actions in order to reach an objective. Formally, this is equivalent to finding a path from an initial state to a goal/marked state in a huge automaton. The latter is specified by a collection of resources, that may be available or not (which defines a state), and actions that consume and produce resources (which defines a transition). In the case of optimal planning, actions have a cost, and the objective is to find a path of minimal cost to the goal.

Our interest in this problem is threefold. First, it is naturally an instance of a concurrent system, given that actions have local effects on resources. Secondly, it is a weak form of an optimal control problem for a concurrent/distributed system. Finally, we are interested in distributed solutions to such problems, which is an active topic in the planning community under the name of "factored planning."

Our previous contribution to the domain was the first optimal factored planning algorithm [47]. The main idea is to represent a planning problem as a network of interacting weighted automata, the objective being to jointly drive all of them to a target state, while minimizing the cost of their joint trajectory. We have developed and tested [53] a distributed algorithm to solve this problem, based on a weighted automata calculus, and that takes the shape of a message passing procedure. Components perform local computations, exchange messages with their neighbors, in an asynchronous manner, and the procedure converges to the path that each component should follow. The optimal global plan is thus given as a tuple of (compatible) local plans, i.e. a partial order of actions.

In 2012, we have extended this framework in two directions. The first one considers large planning problems for which the interaction graph of components is not a tree. It is well known that message passing algorithms (also called belief propagation) is optimal on trees. To recover such a situation where distributed optimal planning can be resolved exactly, one therefore has to smartly group components into larger ones in order to recover a tree of larger components. This is done at the expense of the complexity in the resolution of local planning problems (which augments exponentially with the number of assembled components). Alternately, one can also ignore that the graph is not a tree, and thus use the so-called loopy belief propagation, which requires minor adaptations. This results in a new approach to the resolution of planning problems, where approximate solutions are provided: one can check that the computed plans are valid, but their optimality is not guaranteed. We have experimented this turbo-planning idea on a series of random benchmarks, some of them being not accessible to standard planning methods. The results are surprisingly good: distributed plans are found in most cases, and are often close to optimal. However, no theoretical results can yet support this phenomenon [30].

The second extension to distributed planning concerns the multi-agent version of the central A* (A-star) algorithm, which is at the core of numerous planners. By contrast with the previous setting, we do not build all plans here, in a distributed manner, but perform a search for an optimal plan. The centralized version of A* performs a depth-first search of a winning path in a graph, guided by some heuristic function that orients the search towards the goal. In our setting, several path searches must be performed in the graphs of the different components (or local planning problems), under the constraint that the provided paths are compatible, i.e. agree on the execution of the common actions. The resulting local paths must also be jointly optimal, once their costs are added. We have proposed a complete solution to this problem, called A# (A-sharp) [29]. Our efforts now aim at mixing these ideas with the turbo planning approach.

6.2. Fundamental results and algorithms: communication with messages and scenarios

Participants: Loïc Hélouët, Rouwaida Abdallah, Claude Jard, Blaise Genest, Sundararaman Akshay.

In this paragraph, we collect our fundamental results regarding the models and algorithms we use for communicating systems, and in particular, scenarios.

A major challenge with models communicating with messages (e.g.: scenarios) is to *exhibit good classes of models* allowing users to *specify easily complex distributed systems* while *preserving the decidability* of some key problems, such as diagnosis, equality and intersection. Furthermore, when these problems are decidable for the designed models, the second challenge is to design algorithms to keep the *complexity low enough* to allow *implementation in real cases*.

The first part of our work is the study of Time-Constrained MSC graphs (TC-MSGS for short). Time-constrained MSCs (TC-MSCs) are simply MSCs decorated with constraints on the respective occurrence dates of events. The semantics of a TC-MSC T is a dated MSC, that is a MSC where events are associated with an occurrence date. For a given TC-MSC, there can be an infinite set $L(T)$ of dated MSCs satisfying its constraints. Note however that some time-constraints in a TC-MSC may not be satisfiable, and hence $L(T)$ can simply be empty. TC-MSCs can be extended by composition mechanisms such as TC-MSC graphs. TC-MSC graphs are simply automata labeled by TC-MSC. Each path ρ of a TC-MSC G is associated with a TC-MSC T_ρ obtained by concatenation of TC-MSC along ρ. The language $L(G) = \bigcup_{\rho \text{ path of } G} L(T_\rho)$ of a TC-MSC Graph is then the union of all dated MSCs associated with paths of G. Because of inconsistent timing constraints, some path may have no possible realization (i.e $L(T_\rho = \varnothing)$). One can even design a MSC Graph G such that $L(G) = \varnothing$ - such TC-MSC graph is clearly inconsistent-. It has been shown [49] that checking whether $L(G) = \varnothing$ is an undecidable problem in general, but can be decided for the restricted subclass of regular TC-MSC graphs (that have the expressive power of event-count timed automata). We have proposed two restrictions allowing for the decision of emptiness. The first one is K-drift boundedness, which imposes for a fixed integer K that for every T_ρ there exists one dated realization such that for every pair of events e, f appearing in the same transition of G, the dates of e and f differ by at most K. We have shown that K-drift boundedness is decidable in a symbolic and efficient way, and that for $K-$drift bounded TC-MSC graphs, emptiness is decidable. This extends decidability results beyond regular specifications. The second restriction is K-non-zenoness, which imposes that for a fixed K, for every path ρ of G, there exists one realization such that at every date d, at most K events occur between dates d and $d+1$. When a TC-MSC graph is A-drift-bounded and B-non-zeno, then $L(G)$ has a regular set of representants, which opens the way for more involved model-checking applications [10]. We actually succeeded to use a different technique by symbolically encoding the configuration reached. It allows to remove the K-non-zeno restriction, we don't need the seminal result on timed automata of Alur-Dill 1994, and we have a true partial order algorithm, which does not need to consider different interleavings of the same execution [18].

The second part of our work is the study of realistic implementation of scenarios. The main idea is to propose distributed implementation (communicating state machines) of High-level MSCs that do not contain deadlocks, and behave exactly as the original specification. It is well known [51] that a simple projection of a HMSC on each of its processes to obtain communicating finite state machines results in an implementation with more behaviors than the original specification. An implementation of a HMSC H is considered as consistent if and only if it exhibits the same prefix closed set of behaviors as H. We have proposed an implementation solution that uses local controllers allows the distributed synthesized behavior to remain consistent with the original specification. This work has been implemented in our scenario prototype (see the Software section). This synthesis algorithm is consistent for a particular syntactic class of scenarios, namely the class of local HMSCs. This work was accepted for publication in [14].

6.3. Fundamental results and algorithms: timed models

Participants: Claude Jard, Aurore Junier, Sundararaman Akshay, Loïc Hélouët.

Our work on that subject mainly concerns Time Petri Nets (TPNs) and their robustness. Robustness of timed systems aims at studying whether infinitesimal perturbations in clock values can result in new discrete behaviors. A model is robust if the set of discrete behaviors is preserved under arbitrarily small (but positive) perturbations. We have tackled this problem for Time Petri Nets (TPNs for short) by considering the model of parametric guard enlargement which allows time-intervals constraining the firing of transitions in TPNs to be enlarged by a (positive) parameter.

We have shown that TPNs are not robust in general and that checking if they are robust with respect to standard properties (such as boundedness, safety) is undecidable. We have also provided two decidable robustly bounded subclasses of TPNs, and shown that one can effectively build a timed automaton which is timed bisimilar even in presence of perturbations. This allowed us to apply existing results for timed automata to these TPNs and show further robustness properties. This work was published in [20].

In a second work, we have considered robustness issues in Time Petri Nets (TPN) under constraints imposed by an external architecture. Our main objective was to check whether a timed specification, given as a TPN behaves as expected when subject to additional time and scheduling constraints. These constraints are given by another TPN that constrains the specification via read arcs. Our robustness property says that the constrained net does not exhibit new timed or untimed behaviors. We show that this property is not always guaranteed but that checking for it is always decidable in 1-safe TPNs. We further show that checking if the set of untimed behaviors of the constrained and specification nets are the same is also decidable. Next we turn to the more powerful case of labeled 1-safe TPNs with silent transitions. We show that checking for the robustness property is undecidable even when restricted to 1-safe TPNs with injective labeling, and exhibit a sub-class of 1-safe TPNs (with silent transitions) for which robustness is guaranteed by construction. This sub-class already lies close to the frontiers of intractability. This work was published in [19].

Finally, in cooperation with IRCCyN in Nantes, we defined a more general model, called "clock transition systems", which generalizes both TPNs and networks of timed automata [32]. This model will allow us to transfer new results on TPNs to the timed automata community.

6.4. Fundamental results and algorithms: dynamic epistemic logic

Participants: Guillaume Aucher, François Schwarzentruber.

Within the research line related to Dynamic Epistemic Logic (DEL), we have addressed two parallel lines of research, which have resulted in two publications [22] and [21]. The first deals with the computational complexity of the model checking problem and the satisfiability problem of DEL and the second deals with providing formal means to reason about the effects of sequences of events on the beliefs of multiple agents when these events are only partially specified. This second line of research is a continuation of the work started last year and was motivated by concerns and problems stemming from the Univerself project of Eric Fabre about IMS network.

1. Although DEL is an influential logical framework for representing and reasoning about information change, little is known about the computational complexity of its associated decision problems. In fact, we only know that for public announcement logic, a fragment of DEL, the satisfiability problem and the model-checking problem are respectively PSPACE-complete and in P. We contributed to fill this gap by proving that for the DEL language with event models, the model-checking problem is, surprisingly, PSPACE-complete. Also, we proved that the satisfiability problem is NEXPTIME-complete. In doing so, we provided a sound and complete tableau method deciding the satisfiability problem.
2. Let us consider a sequence of formulas providing partial information about an initial situation, about a set of events occurring sequentially in this situation, and about the resulting situation after the occurrence of each event. From this whole sequence, we want to infer more information, either about the initial situation, or about one of the events, or about the resulting situation after one of the events. Within the framework of Dynamic Epistemic Logic, we show that these different kinds of problems are all reducible to the problem of inferring what holds in the final situation after the occurrence of

all the events. We then provide a tableau method deciding whether this kind of inference is valid. We implement it in LotrecScheme and show that these inference problems are NEXPTIME-complete. We extend our results to the cases where the accessibility relation is serial and reflexive and illustrate them with the coordinated attack problem.

Parallely to the study of abstract dynamic epistemic logic, we initiate the study of the interaction of argumentation theory and epistemic reasoning [33].

6.5. Fundamental results and algorithms: statistical model checking

Participants: Sean Sedwards, Benoit Boyer, Kevin Corre, Cyrille Jégourel, Axel Legay.

Our work on statistical model checking (SMC) avoids an explicit representation of the state space by building a statistical model of the executions of a system and giving results within confidence bounds. The key challenges of this approach are to reduce the length (simulation steps and cpu time) and number of simulation traces necessary to achieve a result with given confidence. Rare properties pose a particular problem in this respect, since they are not only difficult to observe but their probability is difficult to bound. A further goal is to make a tool where the choice of modeling language and logic are flexible.

We have developed the prototype of a compact, modular and efficient SMC platform which we have named *PLASMA* (PLatform for Statistical Model checking Algorithms). PLASMA incorporates an efficient discrete event simulation algorithm and features an importance sampling engine that can reduce the necessary number of simulation runs when properties are rare. We have found that PLASMA performs significantly better than PRISM (the de facto reference probabilistic model checker) when used in a similar mode: PLASMA's simulation algorithm scales with a lower order and can handle much larger models. When using importance sampling, PLASMA's performance with rare properties is even better.

Plasma has been embedded in a tool chain for the design and the verification of Systems of Systems. The tool has also been used in a planing algorithm.

6.6. Fundamental results and algorithms: quantitative model checking and quantitative specification theories

Participants: Ulrich Fahrenberg, Blaise Genest, Axel Legay, Sundararaman Akshay, Louis-Marie Traonouez, Benoit Delahaye.

In 2012 we have successfully widened the applicability of interface and specification theories to systems with quantitative information such as energy usage, time constraints, or hybrid variables. Building on work done in 2011, we have introduced general quantitative specification theories. These provide a framework for reasoning about a wide range of different specification theories for different quantitative settings. We have provide one particularly important instantiation of the framework, which allows quantitative reasoning about real-time specifications.

Work on timed specifications theory has been continued in 2012 around the tool ECDAR. New case studies have been tested using the tool. These results, published in STTT, demonstrate the interest of the compositional approach for analyzing large systems. Besides the theory of robust specifications has been extended to allow a parametric estimation of the robustness. These results have been implemented in a new tool PyECDAR.

In 2012, we also successfully pursued our work on probabilistic specification theories by enhancing the framework of Abstract Probabilistic Automata, that we introduced in 2010, with several new operators. We first introduced a notion of satisfaction for stuttering implementations and showed how this new notion fits in the framework of APAs. Stuttering implementations are Probabilistic Automata that allow "silent" transitions by using local variables that are invisible to the specification. In this context, we also introduced a new logic, called ML-(A)PA that allows specifying properties of APA specifications and stuttering PA implementations. Our next contribution was to introduce a new difference operator. Given two specification APAs, their difference is a new APA that represents all implementations satisfying the one but not the other. This novel operator brings a new light to the well-known domain of counter-example generation.

Concerning Markov Chains, we have developed a new logic, LTL-I, which can only reason about fixed intervals instead of point values. We developed ϵ under and over approximation of formulas of this logics in [17], with associated algorithms. In all but few cases, we know that results of these algorithms are exact answers, while we didn't need to compute precisely and explicitly every probability involved. Another line of research is to consider very large Markov chain represented by Dynamic Bayesian Network. In [15], we compute only approximated results, as the size of the underlying Markov Chain is too big. However, evaluation of the algorithm shows small errors of our algorithm compared with the exact value.

6.7. Specific studies: Web services orchestrations

Participants: Ajay Kattepur, Albert Benveniste, Claude Jard.

Web services *orchestrations* and *choreographies* refer to the composition of several Web services to perform a co-ordinated, typically more complex task. We decided to base our study on a simple and clean formalism for WS orchestrations, namely the ORC formalism proposed by Jayadev Misra and William Cook [55].

Main challenges related to Web services QoS (Quality of Service) include: 1/ To model and quantify the QoS of a service. 2/ To establish a relation between the QoS of queried Web services and that of the orchestration (contract composition); 3/ To monitor and detect the breaching of a QoS contract, possibly leading to a reconfiguration of the orchestration. Typically, the QoS of a service is modeled by a *contract* (or Service Level Agreement, SLA) between the provider and the consumer of a given service. To account for variability and uncertainty in QoS, we proposed in previous work soft probabilistic contracts specified as probabilistic distributions involving the different QoS parameters; we studied *contract composition* for such contracts; we developed probabilistic QoS contract monitoring; and we studied the *monotonicity* of orchestrations; an orchestration is monotonic if, when a called service improves its performance, then so does the overall orchestration.

Last year, in the framework of the Associated Team FOSSA with the University of Texas at Austin (John Thywissen (PhD), Jayadev Misra and William Cook), we extended our approach to general QoS parameters, i.e., beyond response time. We now encompass composite parameters, which are thus only partially, not totally, ordered. We developed a general algebra to capture how QoS parameters are transformed while traversing the orchestration and we extended our study of monotonicity. Finally, we have developed corresponding contract composition procedures. This year, John Thywissen (from UT Austin) and Ajay Kattepur have prototyped a toolbox for Orc to support QoS-management. A journal paper is submitted.

A key task in extending Orc for QoS was to extend the Orc engine so that causalities between the different site calls are made explicit at run time while execution progresses. This benefits from our previous work on Orc semantics, but a new set of rules has been proposed to generate causalities in an efficient way, by covering new features of the language. This is joint work of Claude Jard, Ajay Kattepur and John Thywissen from Austin. An implementation on Orc is under development and a publication is in preparation.

Besides this main line of work, the additional topic of *Negotiation Strategies for Probabilistic Contracts in Web Services Orchestrations* has been addressed by Ajay Kattepur as part of his thesis, see [31]. Service Level Agreements (SLAs) have been proposed in the context of web services to maintain acceptable quality of service (QoS) performance. This is specially crucial for composite service orchestrations that can invoke many atomic services to render functionality. A consequence of SLA management entails efficient negotiation proto- cols among orchestrations and invoked services. In composite services where data and QoS (modeled in a probabilistic setting) interact, it is difficult to pick an individual atomic service to negotiate with. A superior improvement in one negotiated domain (eg. latency) might mean deterioration in another domain (eg. cost). In this work, we propose an integer programming formulation based on first order stochastic dom- inance as a strategy for re-negotiation over multiple services. A consequence of this is better end-to-end performance of the orchestration compared to random strategies for re-negotiation. We also demonstrate this optimal strategy can be applied to negotiation protocols specified in languages such as Orc. Such strategies are necessary for composite services where QoS contributions from individual atomic services vary significantly.

6.8. Specific studies: active documents and web services

Participants: Albert Benveniste, Loïc Hélouët, Sundararaman Akshay.

Active Documents have been introduced by the GEMO team at Inria Futurs, headed by Serge Abiteboul, mainly through the language *Active XML* (or *AXML* for short). AXML is an extension of XML which allows to enrich documents with *service calls* or sc's for short. These sc's point to web services that, when triggered, access other documents; this materialization of sc's produces in turn AXML code that is included in the calling document. One therefore speaks of dynamic or intentional documents. In the past years, we have collaborated with the GEMO team to study a distributed version of their language.

Last year, we have developed a distributed Active XML engine, which can be distributed over a network. We have built a lightweight experimentation platform, made of four Linux machines, that run DAXML services and communicate with one another. This year, we have started an experiment with a case study. We have proposed a distributed chess service palteform; the main idea is to use choreographies to provide solutions for chess problems, relying on an orchestration of specialized services for different phases of a game (opening, end of game, or collecting positions databases. We expect preliminary results in 2013.

Last year, we have proposed a new model, that combines arbitrary numbers of finite workflows, hence allowing for the definition of sessions. Sessions is a central paradigm in web-based systems. As messages exchange between two sites need not follow the same route over the net, a site can not rely on the identity of machines to uniquely define a transaction. This unique identification is essential: a commercial site, for instance, needs to manage several interactions at a given time. The current trend, as in BPEL, is to associate a unique identifier with each session. Modeling realistic sessions hence often forces to include session counters, and hence render most of properties undecidable. The session formalism studied in 2011 can be seen as a mix of BPEL and Orc elements, but was designed to keep several properties decidable (the formalism has the expressive power of reset Petri nets). The strength of this formalism is to allow designing systems that use sessions without the obligation to provide identifiers. Its drawback is that it only allows for the design of systems with a fixed number of agents. This year, we have continued extending last year's work with Ph. Darondeau from the S4 Team, and with M. Mukund from the Chennai Mathematical Institute to allow design of systems with sessions and allowing for an arbitrary number of agents.

6.9. Specific studies: network maintenance

Participants: Eric Fabre, Carole Hounkonnou.

This work represents part of our activities within the research group "High Manageability," supported by the common lab of Alcatel-Lucent Bell Labs (ALBLF) and Inria. It concerns a methodology for the graceful shut down and restart of routers in OSPF networks, one of the core protocols of IP networks. A methodology has been proposed to safely switch off the software layer of a router while still maintaining this router in the forwarding plane: the router still forwards packets, but is not able to adapt its routing table to changes in network conditions or topology. Nevertheless, it is possible to check whether this frozen router is harmless or can cause packet losses, through a centralized or distributed algorithm. And if ever it puts the network at risk, minimal patches can be set up temporarily until the router comes back to normal activity. This avoids running twice a global OSPF update at all nodes (one for shutdown of the equipment, one for restart). This work has been patented in June 2012 jointly with Alcatel-Lucent, and a publication on the topic was accepted at IM'2013.

6.10. Specific studies: network and service diagnosis

Participants: Eric Fabre, Carole Hounkonnou.

This work represents part of our activities within the research group "High Manageability," supported by the common lab of Alcatel-Lucent Bell Labs (ALBLF) and Inria. It is also supported by the UniverSelf EU integrated project, and conducted in cooperation with Orange Labs.

The objective is to develop a framework for the joint diagnosis of networks and of the supported services. We are aiming at a model-based approach, in order to tailor the methods to a given network instance and to follow its evolution. We also aim at active diagnosis methods, that collect and reason on alarms provided by the network, but that can also trigger tests or the collection of new observations in order to refine a current diagnosis.

Since 2011, an important effort was dedicated to a key and difficult part of this approach: the definition of a methodology for self-modeling. This consists in automatically building a model of the monitored system, by instantiating generic network elements. There are several difficulties to address:

- The model must capture several layers, from the physical architecture up to the service architecture and its protocols. As a case-study, we have chosen VoIP services on an IMS network, deployed over a wired IP network.
- The model should be hierarchical, to allow for multiscale reasoning, and to reflect the intrinsic hierarchical nature of the managed network.
- The model should be generic, i.e. obtained by assembling component instances coming from a reduced set of patterns, just like a text is obtained by assembling words.
- The model should be adaptive, to capture the evolving part of the network (e.g. introduction of new elements) but also its intrinsically dynamic nature (e.g. opened/closed connections).
- The model should display the hierarchical dependency of resources, specifically the fact that lower-level resources are assembled to provide a support to a higher level resource or functionality.
- The model should allow progressive discovery and refinement: for a matter of size, it is not possible to first build a model of the complete network and then monitor it; one must adopt an approach where the model is build on-line, and where the construction is guided by the progress of the diagnosis algorithms.

Elements of methodology achieving these goals were proposed in 2011, and further refined in 2012. Besides, we have also worked on the definition of generic Bayesian networks, that could translate into mathematical terms the dependency relations between network resources, in order to reason about them for failure diagnosis. A methodology was then designed to reason on such models. The idea is that one should first consider a subset of network resources (at a given granularity), in order to localize the origin of a given malfunction. The natural start point is the graph of all resources involved in the delivery of the malfunctioning service. As the fault localization is statistical, the model is then progressively expanded to capture more network elements and thus more observations, and thus refine the diagnosis. This model expansion is performed by introducing first the most informative network elements, using information theory criteria. The result is a fault localization algorithm that explores only part of the network, and builds at runtime the necessary part of the model it should use to explain a malfunction [28]. The current efforts aim at extending these ideas to allow for the refinement of the model of some component (multiresolution reasoning).

FOCUS Project-Team

6. New Results

6.1. Service-oriented computing

Participants: Mila Dalla Preda, Ornela Dardha, Maurizio Gabbrielli, Elena Giachino, Claudio Guidi, Jacopo Mauro, Fabrizio Montesi, Davide Sangiorgi.

6.1.1. Primitives

In the context of Service-Oriented Architectures (SOAs), the integration of services is an important aspect that is usually addressed by using specific tools, such as Enterprise Service Bus (ESB). Although widely used these ad-hoc solutions do not exploit the possibility of using a mechanism of interface extension to foster the rapid prototyping and deployment of in-the-middle entities that compose services abstracting from the order in which they exchange messages. We have proposed [29] a framework to perform service integration, based on the extension of service interfaces, capturing a class of service integrators that are decoupled from the services they integrate in an SOA. We also provide a reference implementation for the primitive of service integration by extending the Jolie language, thus allowing for the experimentation with real SOA scenarios. We have shown [30] how our methodology differs from the standard practice with ESB.

6.1.2. Contracts and sessions

Contracts are descriptions of the functionalities offered by a component or a service, and of the way these functionalities may be accessed by clients. A contract may include a description of the component capabilities, place constraints on their usage, as well as declare preferences, entitlements and credentials. When a client wants to use one of the functionalities offered, it engages a dialogue (e.g., a sequence of interactions) with the servers; this is usually called a session.

Contracts specify the expected dialogue in a session and they can be expressed as types, usually called *session types* in this context.

A session type describes communication by specifying type and direction of data exchanged between two parties. When session types and session primitives are added to the syntax of the types and terms of a language, they give rise to additional separate syntactic categories. As a consequence, there may be duplication of efforts in the theory: the proofs of properties must be checked both on ordinary types and on session types. We have shown [32] that this duplication is not necessary, by exhibiting an encoding of (dyadic) session types into ordinary types. Using the encoding, the properties of session types are derived as straightforward corollaries.

We have also studied the problem of handling unexpected or unwanted conditions in sessions, that may change the default execution of distributed communication protocols. We have proposed [14] a global escape mechanism; it can handle such events while preserving compatibility of multiparty conversations. This flexibility enables us to model complex exceptions such as criss-crossing global interactions and error handling for distributed cooperating threads. Guided by multiparty session types, our semantics is proven to provide a termination algorithm for global escapes, as well as further safety properties, such as progress within the session and atomicity of escapes with respect to the subset of involved participants.

6.2. Adaptability and faults

Participants: Mario Bravetti, Elena Giachino, Ivan Lanese, Michael Lienhardt, Jacopo Mauro, Davide Sangiorgi, Gianluigi Zavattaro.

6.2.1. Reversibility

We have continued the study of reversibility started in the past years, aimed at developing programming abstractions for reliable distributed systems. We have shown [39] preliminary results on the interplay between reversibility and compensations, which are a main ingredient in many existing techniques for reliability, in particular long running transactions.

We have then applied [43] our reversibility theory to μOz, a concurrent programming language defined by a stack-based abstract machine, and we make it reversible. This is a first step towards the definition of reversible variants of more complex languages. As additional result we show that the memory overhead due to reversibility is optimal as an order of magnitude.

6.2.2. Primitives for adaptable and evolving components

We study primitives for adaptable and evolving components both in an abstract algebraic setting and in a more concrete setting based on the ABS object-oriented language.

We have defined [13] adaptable processes, a concurrent higher-order calculus where processes have a location, and are sensible to actions of update at runtime. This allows us to express a wide range of evolvability patterns. We have also defined [24] a temporal logic over adaptable processes, with examples of the expressiveness of the logic and of its significance in relation to the calculus of adaptable processes.

A different direction has focused on ABS, a concurrent object-oriented language based on futures for asynchronous method invocations and on object groups for concurrency control. We have given [38] an overview of the architectural aspects of ABS: a feature-driven development workflow, a formal notion of deployment components for specifying environmental constraints, and a dynamic component model that is integrated into the language. We have employed an industrial case study to demonstrate how the various aspects work together in practice. In [40] we have focused on the component model and studied techniques allowing safe dynamic reconfiguration. Our approach adds to ABS: i) output ports to represent variability points, ii) critical sections to control when updates of the software can be made and iii) hierarchy to model locations and distribution. These different notions work together to allow dynamic and safe update of a system.

6.2.3. Reconfigurability in the cloud

The cloud is a relevant application domain for FOCUS. We have considered [35] the problem of deploying and (re)configuring resources in a cloud setting, where interconnected software components and services can be deployed on clusters of heterogeneous (virtual) machines that can be created and connected on-the-fly. We introduce a component model to capture similar scenarii from realistic cloud deployments, and instrument automated planning of day-to-day activities such as software upgrade planning, service deployment, elastic scaling, etc. We formalize the model and characterize the feasibility and complexity of configuration achievability.

6.2.4. Delta-Oriented Programming and Software Product Lines

Delta-oriented programming (DOP) provides a technique for implementing Software Product Lines based on modifications (add, remove, modify) to a core program. Unfortunately, such modifications can introduce errors into a program, especially when type signatures of classes are modified in a non-monotonic fashion. To deal with this problem in we have designed [42] a type system for delta-oriented programs based on row polymorphism. This exercise elucidates the close correspondence between delta-oriented programs and row polymorphism.

In [41] we have studied the notion of conflict for a variant of DOP without features, separating out the notions of hard and soft conflict. Specifically, we have defined a language for this subset of DOP and give a precise, formal definition of these notions. We have then extended the type system in [42] to ensure that the computation of a well-typed product will always succeed and has an unambiguous result.

6.3. Resource Control

Participants: Michele Alberti, Ugo Dal Lago, Marco Gaboardi, Daniel Hirschkoff, Simone Martini, Paolo Parisen Toldin, Giulio Pellitta, Barbara Petit, Davide Sangiorgi, Marco Solieri.

In Focus, we study both foundations and methodologies for controlling the amount of resources programs and processes make use of. The employed techniques mainly come from the world of type theory and proof theory, and as such have been used extensively in the context of sequential computation. Interesting results have been obtained recently indicating that those techniques can be quite useful in the concurrent context too, thus being potentially interesting for CBUS.

During 2012, we have continued our work on intensionally complete techniques for the complexity analysis of functional programs. In [15] a relatively complete type system from which complexity of call-by-name terms has been introduced, while in [25] the same approach is used in a call-by-value setting. The introduced methodology allows us to reduce the problem at hand to the verification of a set of first-order proof-obligations. No information is lost along the reduction.

The interpretation method has been the object of a couple of investigations. On the one hand, we have proved a necessary condition for a first-order program to admit a quasi-interpretation [12]: it must be blind, namely it must be somehow insensible to its argument value, but only sensible to their length. Moreover, we have introduced a new methodology for the complexity analysis of higher-order programs based on an higher-order generalizations of ordinary polynomial interpretations and quasi-interpretations [23].

Among the most foundational works in this area, we should also mention those about invariance results on cost models [16], [22]: we proved that in many different cases, the number of beta-reduction steps is an adequate cost-model for the lambda calculus. These results are potentially useful for complexity analysis, in that they show that a natural and quite intuitive cost model is indeed reasonable, meaning that evaluation can be simulated by finer-grained models of computation within a polynomial overhead.

Finally, some of our works have to do with the semantics of various sorts of lambda calculi with linearity constraints: a non-deterministic extension of the call-by-value lambda calculus, which corresponds to the additive fragment of the linear-algebraic lambda-calculus [36]; a lambda calculus with constructors that decomposes the pattern matching à la ML into some atomic rules [44]; a categorical approach to model the programming language SlPCF that has been conceived in order to program only linear functions between Coherence Spaces [20].

We have also continued our work on techniques for ensuring termination of programs, studying [34] how to transport techniques initially devised for processes onto sequential higher-order languages with imperative features (e.g., λ-calculi with references). The method employed makes it possible to combine term rewriting measure-based techniques for termination of imperative languages with traditional approaches to termination in purely functional languages, such as logical relations.

6.4. Verification of extensional properties

Participants: Mario Bravetti, Daniel Hirschkoff, Cosimo Laneve, Jean-Marie Madiot, Tudor Alexandru Lascu, Davide Sangiorgi, Gianluigi Zavattaro.

Extensional refers to properties that have to do with behavioral descriptions of a system (i.e., how a system looks like from the outside). Examples of such properties include classical functional correctness and deadlock freedom. A substantial amount of the work carried out this year has to do with the transfer of techniques from the area of concurrency theory to the investigation of properties in adaptable systems, obiect-oriented concurrent systems, and systems based on specific synchronization mechanisms.

6.4.1. Adaptability

We mentioned earlier the process calculus of adaptable processes [13] and the related temporal logic [24]. In the same papers, we have addressed the (un)decidability of two safety properties related to error occurrence, and we have explained how the proof techniques in [13] can be extended to prove (un)decidability results for the temporal logic.

6.4.2. Object-orientation

We have considered concurrent object-oriented languages with futures and cooperative scheduling. Verification of deadlock in such systems is a nontrivial task due to the dynamic and unbounded creation of futures. We have introduced [45] a technique to prove deadlock freedom for such systems by translating a concrete program to an abstract version of the program, and then encoding such abstract program into a Petri net. Deadlock can be detected on Petri nets via checking the reachability of a distinct marking: absence of deadlocks in the Petri net constitutes deadlock freedom of the concrete system.

6.4.3. Synchronization primitives

We have investigated [33] the impact of node and communication failures on the decidability and complexity of parametric verification of a formal model of ad hoc networks, in which finite state processes communicate via selective broadcast. We have considered three possible kinds of node failures –intermittence, restart, and crash– and three cases of communication failures –nondeterministic message loss, message loss due to conflicting emissions, and detectable conflicts. Interestingly, we have proved that the considered decision problem (reachability of a control state) is decidable for node intermittence and message loss (either nondeterministic or due to conflicts) while it turns out to be undecidable for node restart/crash, and conflict detection. The conclusion is that verification is decidable only when processes are unaware of the occurrence of a failure.

In another line of work, we have studied the impact of dualities and symmetries in synchronization primitives for message-passing processes [37]. We have shown that in languages where input and output are dualisable (e.g., variants of the π-calculus such as πI and fusion), duality breaks with the addition of ordinary input/output types. We have then considered the minimal symmetrical conservative extension of π-calculus input/output types. We have proved duality properties for it. As example of application of the dualities, we have used this language to relate two encodings of λ-calculus, by Milner and by van Bakel and Vigliotti, syntactically quite different from each other. Thus, results on one encoding can be transferred onto the other one.

6.4.4. Coinduction

Induction is a pervasive tool in Computer Science and Mathematics for defining structures and reasoning on them. Coinduction is the dual of induction, and as such it brings in tools that are quite different from those provided by induction. The best known instance of coinduction is bisimulation, mainly employed to define and prove equalities among potentially infinite objects: processes, streams, non-well-founded sets, and so on. Sangiorgi has completed [48], [51] two comprehensive textbooks on bisimulation and coinduction (in [51], Sangiorgi is an editor, and author of two chapter contributions [49], [47]). The books explain the fundamental concepts and techniques, and the duality with induction. A special emphasis is put on bisimulation as a behavioural equivalence for processes. Thus the books also serve as an introduction to models for expressing processes, and to the associated techniques of operational and algebraic analysis.

6.5. Expressiveness of computational models

Participants: Cosimo Laneve, Maurizio Gabbrielli, Gianluigi Zavattaro.

Expressiveness refers to the study of the expressive power of computational models.

We have studied [46] the expressiveness of an actor-based language similar to the language ABS developed in Hats. We have identified the presence/absence of fields as a relevant feature: the dynamic creation of names in combination with fields gives rise to Turing completeness. On the other hand, restricting to stateless actors gives rise to systems for which properties such as termination are decidable. Such decidability result holds in actors with states when the number of actors is finite and the state is read-only.

Our other study of expressiveness has been made on Constraint Handling Rules (CHR), a committed-choice declarative language originally designed for writing constraint solvers and that is nowadays a general purpose language. The study of CHR is interesting within Focus as this kind of languages, having both constraint solving and concurrency features, allow us to express in a natural way quantitative aspects related to resources. Moreover, constraints may be used to describe dynamic adaptation and evolution of systems. CHR programs consist of multi-headed guarded rules which allow one to rewrite constraints into simpler ones until a solved form is reached. Many empirical evidences suggest that multiple heads augment the expressive power of the language (somehow, it can be considered similar to multiple synchronization gathering n processes simultaneously), however no formal result in this direction had been proved so far. We have proved [18] a number of expressiveness results to support such a claim. First, we have analyzed the Turing completeness of CHR with respect to the underlying constraint theory. If the constraint theory is powerful enough then restricting to single head rules does not affect Turing completeness. On the other hand, differently from the case of the multi-headed language, the single head CHR language is not Turing powerful when the underlying signature (for the constraint theory) does not contain function symbols. Then we have proved that, no matter which constraint theory is considered, under some reasonable assumptions it is not possible to encode the CHR language (with multi-headed rules) into a single headed language while preserving the semantics of the programs. Moreover, under some stronger assumptions, considering an increasing number of atoms in the head of a rule augments the expressive power of the language.

FUN Team

5. New Results

5.1. Routing in FUN

Participants: Nicolas Gouvy, Xu Li, Nathalie Mitton.

Wireless sensor and actuator/robot networks need some routing mechanisms to ensure that data travel the network to the sink with some guarantees. The FUN research group has investigated different geographic routing paradigms. It first has considered a static network in which the routing either enhances the energy cost [22], [10], balances the load over nodes [21], [8] or respects traffic priorities [18].

A more complex routing paradigm has been proposed in [25] for k-anycasting. In k-anycasting, a sensor wants to report event information to any k sinks in the network. This is important to gain in reliability and efficiency in wireless sensor and actor networks. In this paper, we describe KanGuRou, the first position-based energy efficient k-anycast routing which guarantees the packet delivery to k sinks as long as the connected component that contains s also contains sufficient number of sinks. A node s running KanGuRou first computes a tree including k sinks among the M available ones, with weight as low as possible. If this tree has $m \geq 1$ edges originated at node s, s duplicates the message m times and runs m times KanGuRou over a subset of defined sinks. Simulation results show that KanGuRou allows up to 62% of energy saving compared to plain anycasting.

We then assumed that the sink that collects data is actually mobile and travels the network. Sensor nodes need thus to update the position of the sink in a smart fashion in order to limit the overhead generated by this update. In [9], we propose a novel localized Integrated Location Service and Routing (ILSR) scheme, based on the geographic routing protocol GFG, for data communications from sensors to a mobile sink in wireless sensor networks. The objective is to enable each sensor to maintain a slow-varying routing next hop to the sink rather than the precise knowledge of quick-varying sink position. In ILSR, sink updates location to neighboring sensors after or before a link breaks and whenever a link creation is observed. Location update relies on flooding, restricted within necessary area, where sensors experience (next hop) change in GFG routing to the sink. Dedicated location update message is additionally routed to selected nodes for prevention of routing failure. Considering both unpredictable and predictable (controllable) sink mobility, we present two versions. We prove that both of them guarantee delivery in a connected network modeled as unit disk graph. ILSR is the first localized protocol that has this property. We further propose to reduce message cost, without jeopardizing this property, by dynamically controlling the level of location update. A few add-on techniques are as well suggested to enhance the algorithm performance. We compare ILSR with an existing competing algorithm through simulation. It is observed that ILSR generates routes close to shortest paths at dramatically lower (90% lower) message cost.

When the network is composed of mobile sensors that have the faculty to control their mobility, this property can be exploited to enhance routing performance. In [3], we are interested in energy-aware routing algorithms that explicitly take advantage of node mobility to improve energy consumption of computed paths. Mobility is a two-sword edge however. Moving a node may render the network disconnected and results in early termination of information delivery. To mitigate these problems, we propose a family of routing algorithm called CoMNet (Connectivity preservation Mobile routing protocols for actuator and sensor NETworks), that uses local information and modifies the network topology to support resource efficient transmissions. Our extensive simulations show that CoMNet has high energetic performance improvement compared to existing routing algorithms. More importantly, we show that CoMNet guarantees network connectivity and efficient resource consumption.

5.2. Self-organization

Participants: Tony Ducrocq, Xu Li, Nathalie Mitton.

Self-organization encompasses several mechanisms [35]. This year, the FUN research group contributes to some of them such as neighbor discovery, localization, clustering and topology control in FUN.

5.2.1. Neighbor discovery

To perform routing or any specific task, a node needs to discover its neighbors. Hello protocol is the basic technique for neighborhood discovery in wireless ad hoc networks. It requires nodes to claim their existence/aliveness by periodic 'hello' messages. Central to a hello protocol is the determination of hello message transmission rate. No fixed optimal rate exists in the presence of node mobility. The rate should in fact adapt to it, high for high mobility and low for low mobility. In [31], we combine parameters of the neighborhood discovery (sending frequency of hello messages and changes in the neighborhood tables) and transmission range of the nodes. We present two algorithms that adapt transmission range of the sensors in a mobile WSN by still adapting frequency of hello messages in order to save energy and get accurate neighborhood tables. The first solution is based on the knowledge of turnover - change in the number of neighbors in consecutive iterations of the neighborhood discovery - used in conjunction with an adaptation of the message frequency and the transmission range, minimizing overall transmission cost of hello messages. The second solution is based on the computation of optimal range knowing the nodes' speed. Both algorithms are based on theoretical analysis. Results show that they are energy efficient and outperform solutions of the literature by maintaining high accuracy.

5.2.2. Topology control

Topology control is a tool for self-organizing wireless networks locally. It allows a node to consider only a subset of links/neighbors in order to later reduce computing and memory complexity. Topology control in wireless sensor networks is an important issue for scalability and energy efficiency. It is often based on graph reduction performed through the use of Gabriel Graph or Relative Neighborhood Graph. This graph reduction is usually based on geometric values.

In [7], we propose a radically new family of geometric graphs, i.e., Hypocomb, Reduced Hypocomb and Local Hypocomb for topology control. The first two are extracted from a complete graph; the last is extracted from a Unit Disk Graph (UDG). We analytically study their properties including connectivity, planarity and degree bound. All these graphs are connected (provided the original graph is connected) planar. Hypocomb has unbounded degree while Reduced Hypocomb and Local Hypocomb have maximum degree 6 and 8, respectively. To our knowledge, Local Hypocomb is the first strictly-localized, degree-bounded planar graph computed using merely 1-hop neighbor position information. We present a construction algorithm for these graphs and analyze its time complexity. Hypocomb family graphs are promising for wireless ad hoc networking. We report our numerical results on their average degree and their impact on FACE [39] routing. We discuss their potential applications and some open problems.

5.2.3. Localization

In mobile-beacon assisted sensor localization, beacon mobility scheduling aims to determine the best beacon trajectory so that each sensor receives sufficient beacon signals with minimum delay. We propose a novel DeteRministic bEAcon Mobility Scheduling (DREAMS) algorithm [6], without requiring any prior knowledge of the sensory field. In this algorithm, beacon trajectory is defined as the track of depth-first traversal (DFT) of the network graph, thus deterministic. The mobile beacon performs DFT under the instruction of nearby sensors on the fly. It moves from sensor to sensor in an intelligent heuristic manner according to RSS (Received Signal Strength)-based distance measurements. We prove that DREAMS guarantees full localization (every sensor is localized) when the measurements are noise-free. Then we suggest to apply node elimination and topology control (Local Minimum Spanning Tree) to shorten beacon tour and reduce delay. Through simulation we show that DREAMS guarantees full localization even with noisy distance measurements. We evaluate its performance on localization delay and communication overhead in comparison with a previously proposed static path based scheduling method.

5.2.4. Clustering

Clustering in wireless sensor networks is an efficient way to structure and organize the network. It aims to identify a subset of nodes within the network and bind it a leader (i.e. cluster-head). This latter becomes in charge of specific additional tasks like gathering data from all nodes in its cluster and sending them by using a longer range communication to a sink. As a consequence, a cluster-head exhausts its battery more quickly than regular nodes. In [14], we present BLAC, a novel Battery-Level Aware Clustering family of schemes. BLAC considers the battery-level combined with another metric to elect the cluster-head. It comes in four variants. The cluster-head role is taken alternately by each node to balance energy consumption. Due to the local nature of the algorithms, keeping the network stable is easier. BLAC aims to maximize the time with all nodes alive to satisfy application requirements. Simulation results show that BLAC improves the full network lifetime 3-time more than traditional clustering schemes by balancing energy consumption over nodes and still delivering high data percentage.

5.3. Self-deployment

Participants: Milan Erdelj, Xu Li, Karen Miranda, Enrico Natalizio, Tahiry Razafindralambo, Dimitris Zorbas.

Robot self-deployment may have different purposes. The FUN research group has addressed four of them that are *(i)* area coverage, *(ii)* barrier coverage, *(iii)* point of interest coverage and *(iv)* deployment for substitution networks.

5.3.1. Area coverage

In [1], with the focus on the self-organizing capabilities of nodes in WSRN, we propose a movement-assisted technique for nodes self-deployment. Specifically, we propose to use a neural network as a controller for nodes mobility and a genetic algorithm for the training of the neural network through reinforcement learning. This kind of scheme is extremely adaptive, since it can be easily modified in order to consider different objectives and QoS parameters. In fact, it is sufficient to consider a different kind of input for the neural network to aim for a different objective. All things considered, we propose a new method for programming a WSRN and we show practically how the technique works, when the coverage of the network is the QoS parameter to optimize. Simulation results show the flexibility and effectiveness of this approach even when the application scenario changes (e.g., by introducing physical obstacles).

In [4], we tackle the issue in a different way. We leverage prediction by exploiting temporal-spatial correlations among sensory data. The basic idea lies in that a sensor node can be turned off safely when its sensory information can be inferred through some prediction methods, like Bayesian inference. We adopt the concept of entropy in information theory to evaluate the information uncertainty about the region of interest (RoI). We formulate the problem as a minimum weight sub-modular set cover problem, which is known to be NP hard. To address this problem, an efficient centralized truncated greedy algorithm (TGA) is proposed. We prove the performance guarantee of TGA in terms of the ratio of aggregate weight obtained by TGA to that by the optimal algorithm. Considering the decentralization nature of WSNs, we further present a distributed version of TGA, denoted as DTGA, which can obtain the same solution as TGA. The implementation issues such as network connectivity and communication cost are extensively discussed. We perform real data experiments as well as simulations to demonstrate the advantage of DTGA over the only existing competing algorithm and the impacts of different parameters associated with data correlations on the network lifetime.

In [34], [13], we leverage some assumptions. One of the main operations in wireless sensor networks is the surveillance of a set of events (targets) that occur in the field. In practice, a node monitors an event accurately when it is located closer to it, while the opposite happens when the node is moving away from the target. This detection accuracy can be represented by a probabilistic distribution. Since the network nodes are usually randomly deployed, some of the events are monitored by a few nodes and others by many nodes. In applications where there is a need of a full coverage and of a minimum allowed detection accuracy, a single node may not be able to sufficiently cover an event by itself. In this case, two or more nodes are needed to collaborate and to cover a single target. Moreover, all the nodes must be connected with a base station that collects the monitoring data. In this paper we describe the problem of the minimum sampling quality, where an event must be sufficiently detected by the maximum possible amount of time. Since the probability of detecting a single target using randomly deployed static nodes is quite low, we present a localized algorithm based on mobile nodes. Our algorithm sacrifices a part of the energy of the nodes by moving them to a new location in order to satisfy the desired detection accuracy. It divides the monitoring process in rounds to extend the network lifetime, while it ensures connectivity with the base station. Furthermore, since the network lifetime is strongly related to the number of rounds, we propose two redeployment schemes that enhance the performance of our approach by balancing the number of sensors between densely covered areas and areas that are poorly covered. Finally, our evaluation results show an over 10 times improvement on the network lifetime compared to the case where the sensors are static. Our approaches, also, outperform a virtual forces algorithm when connectivity with the base station is required. The redeployment schemes present a good balance between network lifetime and convergence time.

5.3.2. Barrier coverage

Barrier coverage problem in emerging mobile sensor networks has been an interesting research issue. Existing solutions to this problem aim to decide one-time movement for individual sensors to construct as many barriers as possible, which may not work well when there are no sufficient sensors to form a single barrier. In [19], we try to achieve barrier coverage in sensor scarcity case by dynamic sensor patrolling. In specific, we design a periodic monitoring scheduling (PMS) algorithm in which each point along the barrier line is monitored periodically by mobile sensors. Based on the insight from PMS, we then propose a coordinated sensor patrolling (CSP) algorithm to further improve the barrier coverage, where each sensor's current movement strategy is decided based on the past intruder arrival information. By jointly exploiting sensor mobility and intruder arrival information, CSP is able to significantly enhance barrier coverage.We prove that the total distance that the sensors move during each time slot in CSP is the minimum. Considering the decentralized nature of mobile sensor networks, we further introduce two distributed versions of CSP: S-DCSP and G-DCSP. Through extensive simulations, we demonstrate that CSP has a desired barrier coverage performance and S-DCSP and G-DCSP have similar performance as that of CSP.

5.3.3. Point of Interest coverage

The coverage of Points of Interest (PoI) is a classical requirement in mobile wireless sensor applications. Optimizing the sensors self-deployment over a PoI while maintaining the connectivity between the sensors and the base station is thus a fundamental issue. This algorithm addresses the problem of autonomous deployment of mobile sensors that need to cover a predefined PoI with a connectivity constraint. In our algorithm [2], each sensor moves toward a PoI but has also to maintain the connectivity with a subset of its neighboring sensors that are part of the Relative Neighborhood Graph (RNG). The Relative Neighborhood Graph reduction is chosen so that global connectivity can be provided locally. Our deployment scheme minimizes the number of sensors used for connectivity thus increasing the number of monitoring sensors. Analytical results, simulation results and practical implementation are provided to show the efficiency of our algorithm.

We then extended this coverage to multiple points of interest in [15], [16]. Indeed, the problems of multiple PoI coverage, environment exploration and data report are still solved separately and there are no works that combine the aforementioned problems into a single deployment scheme. In this work, we have extended [2] to multiple PoI coverage and combined it to and environment exploration in order to capture the dynamics of the monitored area. We examine the performance of our scheme through extensive simulation campaigns.

5.3.4. Substitution networks

A substitution network is a temporary network that will be deployed to support a base network in trouble and help it to provide best service. [11], [24] present how the mobility of routers impacts the performance of a wireless substitution network. To that end, we simulate a scenario where a wireless router moves between three static nodes, a source and two destinations of UDP traffic. Specifically, our goal is to deploy or redeploy the mobile relays so that application-level requirements, such as data delivery or latency, are met. Our proposal for a mobile relay achieves these goals by using an adaptive approach to self-adjust their position based on local information. We obtain results on the performance of end-to-end delay, jitter, loss percentage, and throughput under such mobility pattern for the mobile relay. We show how the proposed solution is able to adapt to topology changes and to the evolution of the network characteristics through the usage of limited neighborhood knowledge.

5.4. MAC layer

Participant: Tahiry Razafindralambo.

Multihop wireless networks are used to provide Internet connectivity to the users and the level of performance and quality expected by these users are increasing. In order to meet these performance and quality requirements, wireless communications should be enhanced. Previous works from the literature show that the performance and quality provided by an IEEE 802.11-based multihop wireless network are far from optimal and that there exist different ways to increase the efficiency and the quality of service of such a network. Some studies show that using the medium state as a parameter to tune the behavior of an IEEE 802.11-based multihop network is an appropriate way to proceed. A station in a IEEE 802.11-based multihop wireless network senses the medium either busy or idle. The durations of idle periods and busy periods and their distributions have a clear impact on the network and nodes performance. The understanding of the relationship between these indicators, namely idle and busy periods, the network topology and the traffic, would give new insights to enhance the performance and quality of multihop wireless networks. Due to its multihop and distributed nature, the characterization of idle period durations is difficult in such a network. In [27], [26], we explore the characterization of idle period distribution by proposing a new analytical model and provides an application of this characterization with the design of an adaptive backoff algorithm based on idle periods.

5.5. Servicing

Participants: Xu Li, Kalypso Magklara, Nathalie Mitton, Tahiry Razafindralambo, Dimitris Zorbas.

Servicing wireless sensor networks include many primitives. It can range from cloud connection [12] to mobile IPv6 management [29] going through energy prediction [20] and launching mobile robots on request of a specific demand [5] or to reload sensors [23], [17].

5.5.1. Node reloading

A critical problem of wireless sensor networks is the network lifetime, due to the device's limited battery lifetime. The nodes are randomly deployed in the field and the system has no previous knowledge of their position. To tackle this problem, in [23], we use a mobile robot, that discovers the nodes around it and replaces the active nodes, whose energy is drained, by fully charged inactive nodes. We propose two localized algorithms, that can run on the robot and that decide, which nodes to replace. We simulate our algorithms and our findings show that all nodes that fail are replaced in a short period of time.

In [17] we focus on an emerging kind of cooperative networking system in which a small team of robotic agents lies at a base station. Their mission is to service an already-deployed WSN by periodically replacing all damaged sensors in the field with passive, spare ones so as to preserve the existing network coverage. This novel application scenario is here baptized as "multiple-carrier coverage repair" (MC2R) and modeled as a new generalization of the vehicle routing problem. A hybrid metaheuristic algorithm is put forward to derive nearly-optimal sensor replacement trajectories for the robotic fleet in a short running time. The composite scheme relies on a swarm of artificial fireflies in which each individual follows the exploratory principles featured by Harmony Search. Infeasible candidate solutions are gradually driven into feasibility under the influence of a weak Pareto dominance relationship. A repair heuristic is finally applied to yield a full-blown solution. To the best of our knowledge, our scheme is the first one in literature that tackles MC2R instances. Empirical results indicate that promising solutions can be achieved in a limited time span.

5.5.2. Energy prediction

One way to improve energy supply for sensor nodes is through ambient energy harvesting from solar, thermal or vibration energy sources coupled with rechargeable energy storage. Wireless sensors have to adapt to the stochastic nature of the energy harvesting sources. We are convinced that predicting the temporal availability of ambient energy resources is vital to plan the harvesting efficiency, optimum resource utilization and energy conservation within sensor nodes. In [20] we propose a novel two stage Autoregressive Weather conditioned Solar Energy Prediction (AWSEP) model which is characterized by low computational complexity and is used to accurately estimate the amount of solar energy that will be harvested in the near future in a particular region. Our algorithm re-learns the model parameters during the prediction processing situations where the prediction error becomes larger than a predefined prediction error threshold mainly because of the unreliable nature of outdoor solar energy sources caused by changes in weather conditions. The proposed AWSEP model performance is evaluated by varying energy harvesting source prediction intervals, sampling rates, trade-offs in prediction accuracy and computational costs using real solar datasets. We concluded that AWSEP algorithm is more accurate, has reduced computational complexity and memory utilization than other prediction schemes in literature. Our proposed algorithm can assist a node to automatically adapt to the changing weather conditions for effective power management and sensing task scheduling.

5.5.3. Servicing sensor nodes

Due to the robots' potential to unleash a wider set of networking means and thus augment the network performance, WSRNs have rapidly become a hot research area. In [5], we elaborate on WSRNs from two unique standpoints: robot task allocation and robot task fulfillment. The former deals with robots cooperatively deciding on the set of tasks to be individually carried out to achieve a desired goal; the latter enables robots to fulfill the assigned tasks through intelligent mobility scheduling.

5.6. Experimenting

Participants: Nathalie Mitton, Julien Vandaele.

One of the goal of the FUN research group is to validate through experimentations and to provide tools for this purpose. Therefore, the FIT platform is deployed, together with a set of tutorials [37]. Nevertheless, we are aware that using testbed platforms for validation is already a great step but it can not satisfy all needs. This is why we also investigate alternatives as emulation. In [28], [32] for instance, we propose a specifically designed experimental setup using a relatively small number of nodes forming a real one-hop neighborhood used to emulate any real WSN. The source node is a fixed sensor, and all other sensors are candidate forwarding neighbors towards a virtual destination. The source node achieves one forwarding step, then the virtual destination position and neighborhood are adjusted. The same source is used again to repeat the process. The main novelty is to spread available nodes regularly following a hexagonal pattern around the central node, used as the source, and selectively use subsets of the surrounding nodes at each step of the routing process to provide the desired density and achieve changes in configurations. Compared to real testbeds, our proposition has the advantages of emulating networks with any desired node distribution and densities, which may not be possible in a small scale implementation, and of unbounded scalability since we can emulate networks with an arbitrary number of nodes. Finally, our approach can emulate networks of various shapes, possibly with holes and obstacles. It can also emulate recovery mode in geographic routing, which appears impossible with any existing approach.

5.7. RFID middleware

Participants: Roberto Quilez, Nathalie Mitton.

The Object Naming System (ONS) is a central lookup service used in the EPCglobal network for retrieving location information about a specific Electronic Product Code (EPC). This centralized solution lacks scalability and fault tolerance and encounters some political issues. In [30], we present the design principles of a fully-distributed multi-root solution for ONS lookup service. In distributed systems, the problem of providing a scalable location service requires a dynamic mechanism to associate identification and location. We design, prototype, and evaluate PRONS, a DHT-based solution for the multi-root problem. We show that PRONS achieves significant performance levels while respecting a number of neutrality requirements.

5.8. VANET

Participants: Enrico Natalizio, Thierry Delot.

Today, thanks to vehicular networks, drivers may receive useful information produced or relayed by neighboring sensors or vehicles (e.g., the location of an available parking space, of a traffic congestion, etc.). In [33], we address the problem of providing assistance to the driver when no recent information has been received on his/her vehicle. Therefore, we present a cooperative scheme to aggregate, store and exchange these events in order to have an history of past events. This scheme is based on a dedicated spatio-temporal aggregation structure using Flajolet-Martin sketches and deployed on each vehicle. Contrary to existing approaches considering data aggregation in vehicular networks, our main goal here is not to save network bandwidth but rather to extract useful knowledge from previous observations. In this paper, we present our aggregation data structure, the associated exchange protocol and a set of experiments showing the effectiveness of our proposal.

In [36], we present a novel vehicular communication protocol, which aims to reduce the effect of broadcast storm problem in VANETs (Vehcular AdHoc NETworks). When the traffic density is above a certain value (e.g., when vehicles are in congested traffic scenarios), one of the most serious problems is the increase of packet collisions and medium contentions among vehicles which attempt to communicate. Our proposed technique, namely Selective Reliable Broadcast protocol (SRB), is intended to limit the number of packet transmissions, by means of opportunistically selecting neighboring nodes, acting as relay nodes. As a result, the number of forwarder vehicles is strongly reduced, while network performance is preserved. SRB belongs to the class of broadcast protocols, and exploits the traditional vehicular partitioning behavior to select forwarders. Each cluster is automatically detected as a zone of interest, whenever a vehicle is approaching, and packets will be forwarded only to selected vehicles, opportunistically elected as cluster-heads. In respect of traditional broadcast approaches, the main strengths of SRB are the efficiency of detecting clusters and selecting forwarders in a fast way, in order to limit the broadcast storm problem. Simulation results have been carried out both in urban and highway scenarios, in order to validate the effectiveness of SRB, in terms of cluster detection and reduction of number of selected forwarders.

GANG Project-Team

4. New Results

4.1. Understanding graph representations

4.1.1. Notions of Connectivity in Overlay Networks

Participants: Yuval Emek, Pierre Fraigniaud, Amos Korman, Shay Kutten, David Peleg.

How well connected is the network? This is one of the most fundamental questions one would ask when facing the challenge of designing a communication network. Three major notions of connectivity have been considered in the literature, but in the context of traditional (single-layer) networks, they turn out to be equivalent. The paper [17], introduces a model for studying the three notions of connectivity in multi-layer networks. Using this model, it is easy to demonstrate that in multi-layer networks the three notions may differ dramatically. Unfortunately, in contrast to the single-layer case, where the values of the three connectivity notions can be computed efficiently, it has been recently shown in the context of WDM networks (results that can be easily translated to our model) that the values of two of these notions of connectivity are hard to compute or even approximate in multi-layer networks. The current paper shed some positive light into the multi-layer connectivity topic: we show that the value of the third connectivity notion can be computed in polynomial time and develop an approximation for the construction of well connected overlay networks.

4.1.2. Connected graph searching

Participants: Lali Barrière, Paola Flocchini, Fedor V. Fomin, Pierre Fraigniaud, Nicolas Nisse, Nicola Santoro, Dimitrios M. Thilikos.

In the graph searching game the opponents are a set of searchers and a fugitive in a graph. The searchers try to capture the fugitive by applying some sequence of moves that include placement, removal, or sliding of a searcher along an edge. The fugitive tries to avoid capture by moving along unguarded paths. The search number of a graph is the minimum number of searchers required to guarantee the capture of the fugitive. In [2], we initiate the study of this game under the natural restriction of connectivity where we demand that in each step of the search the locations of the graph that are clean (i.e. non-accessible to the fugitive) remain connected. We give evidence that many of the standard mathematical tools used so far in classic graph searching fail under the connectivity requirement. We also settle the question on “the price of connectivity”, that is, how many searchers more are required for searching a graph when the connectivity demand is imposed. We make estimations of the price of connectivity on general graphs and we provide tight bounds for the case of trees. In particular, for an n-vertex graph the ratio between the connected searching number and the non-connected one is while for trees this ratio is always at most 2. We also conjecture that this constant-ratio upper bound for trees holds also for all graphs. Our combinatorial results imply a complete characterization of connected graph searching on trees. It is based on a forbidden-graph characterization of the connected search number. We prove that the connected search game is monotone for trees, i.e. restricting search strategies to only those where the clean territories increase monotonically does not require more searchers. A consequence of our results is that the connected search number can be computed in polynomial time on trees, moreover, we show how to make this algorithm distributed. Finally, we reveal connections of this parameter to other invariants on trees such as the Horton–Strahler number.

4.1.3. Computing with Large Populations Using Interactions

Participants: Olivier Bournez, Pierre Fraigniaud, Xavier Koegler.

We define in [12], a general model capturing the behavior of a population of anonymous agents that interact in pairs. This model captures some of the main features of opportunistic networks, in which nodes (such as the ones of a mobile ad hoc networks) meet sporadically. For its reminiscence to Population Protocol, we call our model Large-Population Protocol, or LPP. We are interested in the design of LPPs enforcing, for every $\nu \in [0, 1]$, a proportion ν of the agents to be in a specific subset of marked states, when the size of the population grows to infinity; In which case, we say that the protocol computes ν. We prove that, for every $\nu \in [0, 1]$, ν is computable by a LPP if and only if ν is algebraic. Our positive result is constructive. That is, we show how to construct, for every algebraic number $\nu \in [0, 1]$, a protocol which computes ν.

4.1.4. Collaborative Search on the Plane without Communication

Participants: Ofer Feinerman, Zvi Lotker, Amos Korman, Jean-Sébastien Sereni.

In [19], we use distributed computing tools to provide a new perspective on the behavior of cooperative biological ensembles. We introduce the Ants Nearby Treasure Search (ANTS) problem, a generalization of the classical cow-path problem which is relevant for collective foraging in animal groups. In the ANTS problem, k identical (probabilistic) agents, initially placed at some central location, collectively search for a treasure in the two-dimensional plane. The treasure is placed at a target location by an adversary and the goal is to find it as fast as possible as a function of both k and D, where D is the distance between the central location and the target. This is biologically motivated by cooperative, central place foraging, such as performed by ants around their nest. In this type of search there is a strong preference to locate nearby food sources before those that are further away. We focus on trying to find what can be achieved if communication is limited or altogether absent. Indeed, to avoid overlaps agents must be highly dispersed making communication difficult. Furthermore, if the agents do not commence the search in synchrony, then even initial communication is problematic. This holds, in particular, with respect to the question of whether the agents can communicate and conclude their total number, k. It turns out that the knowledge of k by the individual agents is crucial for performance. Indeed, it is a straightforward observation that the time required for finding the treasure is $\Omega(D + D^2/k)$, and we show in this paper that this bound can be matched if the agents have knowledge of k up to some constant approximation. We present a tight bound for the competitive penalty that must be paid, in the running time, if the agents have no information about k. Specifically, this bound is slightly more than logarithmic in the number of agents. In addition, we give a lower bound for the setting in which the agents are given some estimation of k. Informally, our results imply that the agents can potentially perform well without any knowledge of their total number k, however, to further improve, they must use some information regarding k. Finally, we propose a uniform algorithm that is both efficient and extremely simple, suggesting its relevance for actual biological scenarios.

4.1.5. Memory Lower Bounds for Randomized Collaborative Search and Implications for Biology

Participants: Ofer Feinerman, Amos Korman.

Initial knowledge regarding group size can be crucial for collective performance. We study in [18], this relation in the context of the Ants Nearby Treasure Search (ANTS) problem, which models natural cooperative foraging behavior such as that performed by ants around their nest. In this problem, k (probabilistic) agents, initially placed at some central location, collectively search for a treasure on the two-dimensional grid. The treasure is placed at a target location by an adversary and the goal is to find it as fast as possible as a function of both k and D, where D is the (unknown) distance between the central location and the target. It is easy to see that $T = \Omega(D + D^2/k)$ time units are necessary for finding the treasure. Recently, it has been established that $O(T)$ time is sufficient if the agents know their total number k (or a constant approximation of it), and enough memory bits are available at their disposal. In this paper, we establish lower bounds on the agent memory size required for achieving certain running time performances. To the best our knowledge, these bounds are the first non-trivial lower bounds for the memory size of probabilistic searchers. For example, for every given positive constant ϵ, terminating the search by time $O(\log^{1-\epsilon} k \cdot T)$ requires agents to use $\Omega(\log \log k)$ memory bits.

From a high level perspective, we illustrate how methods from distributed computing can be useful in generating lower bounds for cooperative biological ensembles. Indeed, if experiments that comply with our setting reveal that the ants' search is time efficient, then our theoretical lower bounds can provide some insight on the memory they use for this task.

4.1.6. What Can be Computed without Communications?

Participants: Heger Arfaoui, Pierre Fraigniaud.

When playing the boolean game (δ, f), two players, upon reception of respective inputs x and y, must respectively output a and b satisfying $\delta(a, b) = f(x, y)$, in absence of any communication. It is known that, for $\delta(a, b) = a \oplus b$, the ability for the players to use entangled quantum bits (qbits) helps. In [10], we show that, for δ different from the exclusive-or operator, quantum correlations do not help. This result is an invitation to revisit the theory of dis- tributed checking, a.k.a. distributed verification, currently sticked to the usage of decision functions δ based on the and-operator, hence potentially preventing us from using the potential benefit of quantum effects.

4.1.7. Decidability Classes for Mobile Agents Computing modularity

Participants: Andrzej Pelc, Pierre Fraigniaud.

We establish in [21], a classification of decision problems that are to be solved by mobile agents operating in unlabeled graphs, using a deterministic protocol. The classification is with respect to the ability of a team of agents to solve the problem, possibly with the aid of additional information. In particular, our focus is on studying differences between the decidability of a decision problem by agents and its verifiability when a certificate for a positive answer is provided to the agents. Our main result shows that there exists a natural complete problem for mobile agent verification. We also show that, for a single agent, three natural oracles yield a strictly increasing chain of relative decidability classes.

4.1.8. Randomized Distributed Decision

Participants: Pierre Fraigniaud, Amos Korman, Merav Parter, David Peleg.

The paper [20] tackles the power of randomization in the context of locality by analyzing the ability to "boost" the success probability of deciding a distributed language. The main outcome of this analysis is that the distributed computing setting contrasts significantly with the sequential one as far as randomization is concerned. Indeed, we prove that in some cases, the ability to increase the success probability for deciding distributed languages is rather limited.

We focus on the notion of a $(p, q)-$decider for a language L , which is a distributed randomized algorithm that accepts instances in L with probability at least p and rejects instances outside of L with probability at least q. It is known that every hereditary language that can be decided in t rounds by a $(p, q)-$decider, where $p^2 + q > 1$, can be decided deterministically in $O(t)$ rounds. One of our results gives evidence supporting the conjecture that the above statement holds for all distributed languages and not only for hereditary ones, by proving the conjecture for the restricted case of path topologies. For the range below the aforementioned threshold, namely, $p^2 + q \leq 1$, we study the class $B_k(t)$ (for $k \in \mathbb{N}^* \cup \{\infty\}$) of all languages decidable in at most t rounds by a $(p, q)-$decider, where $p^{1+\frac{1}{k}} + q > 1$. Since every language is decidable (in zero rounds) by a $(p, q)-$decider satisfying $p + q = 1$, the hierarchy B_k provides a spectrum of complexity classes between determinism ($k = 1$, under the above conjecture) and complete randomization ($k = \infty$). We prove that all these classes are separated, in a strong sense: for every integer $k \geq 1$, there exists a language L satisfying $L \in B_{k+1}(0)$ but $L \notin B_k(t)$ for any $t = o(n)$. In addition, we show that $B_\infty(t)$ does not contain all languages, for any $t = o(n)$. In other words, we obtain the hierarchy $B_(t) \subset B_2(t) \subset \cdots \subset B_\infty(t) \subset$ All. Finally, we show that if the inputs can be restricted in certain ways, then the ability to boost the success probability becomes almost null, and in particular, derandomization is not possible even beyond the threshold $p^2 + q = 1$.

4.1.9. The Worst Case Behavior of Randomized Gossip

Participants: Hervé Baumann, Pierre Fraigniaud, Hovhannes A. Harutyunyan, Rémi de Verclos.

In [11] we consider the quasi-random rumor spreading model introduced by Doerr, Friedrich, and Sauerwald in [SODA 2008], hereafter referred to as the list-based model. Each node is provided with a cyclic list of all its neighbors, chooses a random position in its list, and from then on calls its neighbors in the order of the list. This model is known to perform asymptotically at least as well as the random phone-call model, for many network classes. Motivated by potential applications of the list-based model to live streaming, we are interested in its worst case behavior.

Our first main result is the design of an $O(m + nlogn)-$time algorithm that, given any n-node m-edge network G, and any source-target pair $s, t \in V(G)$, computes the maximum number of rounds it may take for a rumor to be broadcast from s to t in G, in the list-based model. This algorithm yields an $O(n(m + nlogn))-$time algorithm that, given any network G, computes the maximum number of rounds it may take for a rumor to be broadcast from any source to any target, in the list-based model. Hence, the list-based model is computationally easy to tackle in its basic version.

The situation is radically different when one is considering variants of the model in which nodes are aware of the status of their neighbors, i.e., are aware of whether or not they have already received the rumor, at any point in time. Indeed, our second main result states that, unless P=NP , the worst case behavior of the list-based model with the additional feature that every node is perpetually aware of which of its neighbors have already received the rumor cannot be approximated in polynomial time within a $\left(\frac{1}{n}\right)^{\frac{1}{2}-\epsilon}$ multiplicative factor, for any $\epsilon > 0$. As a byproduct of this latter result, we can show that, unless P=NP , there are no PTAS enabling to approximate the worst case behavior of the list-based model, whenever every node perpetually keeps track of the subset of its neighbors which have sent the rumor to it so far.

4.1.10. Asymptotic modularity

Participants: Fabien de Montgolfier, Mauricio Soto, Laurent Viennot.

Modularity (Newman-Girvan) has been introduced as a quality measure for graph partitioning. It has received considerable attention in several disciplines, especially complex systems. In order to better understand this measure from a graph theoretical point of view, we study the modularity of a variety of graph classes. In [23], we first consider simple graph classes such as tori and hypercubes. We show that these regular graph families have asymptotic modularity 1 (that is the maximum possible). We extend this result to trees with bounded degree, allowing us to give a lower bound of 2 over average degree for graph classes with low maximum degree (included power law graphs for a sufficiently large exponent).

4.1.11. Modeling social networks

Participants: Nidhi Hegde, Laurent Massoulié, Laurent Viennot.

Social networks offer users new means of accessing information, essentially relying on “social filtering”, i.e. propagation and filtering of information by social contacts. The sheer amount of data flowing in these networks, combined with the limited budget of attention of each user, makes it difficult to ensure that social filtering brings relevant content to the interested users. Our motivation in [26] is to measure to what extent self-organization of the social network results in efficient social filtering. To this end we introduce flow games, a simple abstraction that models network formation under selfish user dynamics, featuring user-specific interests and budget of attention. In the context of homogeneous user interests, we show that selfish dynamics converge to a stable network structure (namely a pure Nash equilibrium) with close-to-optimal information dissemination. We show in contrast, for the more realistic case of heterogeneous interests, that convergence, if it occurs, may lead to information dissemination that can be arbitrarily inefficient, as captured by an unbounded “price of anarchy”. Nevertheless the situation differs when users’ interests exhibit a particular structure, captured by a metric space with low doubling dimension. In that case, natural autonomous dynamics converge to a stable configuration. Moreover, users obtain all the information of interest to them in the corresponding dissemination, provided their budget of attention is logarithmic in the size of their interest set.

4.1.12. Additive Spanners and Distance and Routing Labeling Schemes for Hyperbolic Graphs

Participants: Victor Chepoi, Feodor Dragan, Bertrand Estellon, Michel Habib, Yann Vaxès, Yang Xiang.

δ-Hyperbolic metric spaces have been defined by M. Gromov in 1987 via a simple 4-point condition: for any four points u, v, w, x, the two larger of the distance sums $d(u,v) + d(w,x)$, $d(u,w) + d(v,x)$, $d(u,x) + d(v,w)$ differ by at most 2δ. They play an important role in geometric group theory, geometry of negatively curved spaces, and have recently become of interest in several domains of computer science, including algorithms and networking. In [5], we study unweighted δ-hyperbolic graphs. Using the Layering Partition technique, we show that every n-vertex δ-hyperbolic graph with $\delta \geq 1/2$ has an additive $O(\delta \log n)$-spanner with at most $O(\delta n)$ edges and provide a simpler, in our opinion, and faster construction of distance approximating trees of δ-hyperbolic graphs with an additive error $O(\delta \log n)$. The construction of our tree takes only linear time in the size of the input graph. As a consequence, we show that the family of n-vertex δ-hyperbolic graphs with $\delta \geq 1/2$ admits a routing labeling scheme with $O(\delta \log^2 n)$ bit labels, $O(\delta \log n)$ additive stretch and $O(\log_2(4\delta))$ time routing protocol, and a distance labeling scheme with $O(\log^2 n)$ bit labels, $O(\delta \log n)$ additive error and constant time distance decoder.

4.1.13. Constructing a Minimum phylogenetic Network from a Dense triplet Set

Participants: Michel Habib, Thu-Hien To.

For a given set $\mathcal{L}$ of species and a set $\mathcal{T}$ of triplets on $\mathcal{L}$, we seek to construct a phylogenetic network which is consistent with $\mathcal{T}$ i.e. which represents all triplets of $\mathcal{T}$. The level of a network is defined as the maximum number of hybrid vertices in its biconnected components. When $\mathcal{T}$ is dense, there exist polynomial time algorithms to construct level-$0, 1$ and 2 networks (Aho et al., 1981; Jansson, Nguyen and Sung, 2006; Jansson and Sung, 2006; Iersel et al., 2009). For higher levels, partial answers were obtained in the paper by Iersel and Kelk (2008), with a polynomial time algorithm for simple networks. In [9] this paper, we detail the first complete answer for the general case, solving a problem proposed in Jansson and Sung (2006) and Iersel et al. (2009). For any k fixed, it is possible to construct a level-k network having the minimum number of hybrid vertices and consistent with $\mathcal{T}$, if there is any, in time $O\left(|(T)|^{k+1} n^{\lfloor \frac{4k}{3} \rfloor}\right)$.

4.1.14. Algorithms for Some $H-$Join Decompositions

Participants: Michel Habib, Antoine Mamcarz, Fabien de Montgolfier.

A homogeneous pair (also known as a 2-module) of a graph is a pair $\{M_1, M_2\}$ of disjoint vertex subsets such that for every vertex $x \notin (M_1 \cup M_2)$ and $i \in \{1,2\}$, x is either adjacent to all vertices in M_i or to none of them. First used in the context of perfect graphs [Chvátal and Sbihi 1987], it is a generalization of splits (a.k.a 1-joins) and of modules. The algorithmics to compute them appears quite involved. In [22], we describe an $O(mn^2)-$time algorithm computing (if any) a homogeneous pair, which not only improves a previous bound of $O(mn^3)$ [Everett, Klein and Reed 1997], but also uses a nice structural property of homogenous pairs. Our result can be extended to compute the whole homogeneous pair decomposition tree, within the same complexity. Using similar ideas, we present an $O(nm^2)-$time algorithm to compute a $N-$join decomposition of a graph, improving a previous $O(n^6)$ algorithm [Feder et al. 2005]. These two decompositions are special case of $H-$joins [Bui-Xuan, Telle and Vatshelle 2010] to which our techniques apply.

4.1.15. Detecting 2-joins faster

Participants: Pierre Charbit, Michel Habib, Nicolas Trotignon, Kristina Vušković.

2-joins are edge cutsets that naturally appear in the decomposition of several classes of graphs closed under taking induced subgraphs, such as balanced bipartite graphs, even-hole-free graphs, perfect graphs and claw-free graphs. Their detection is needed in several algorithms, and is the slowest step for some of them. The classical method to detect a 2-join takes $O(n^3m)$ time where n is the number of vertices of the input graph and m the number of its edges. To detect *non-path* 2-joins (special kinds of 2-joins that are needed in all of the known algorithms that use 2-joins), the fastest known method takes time $O(n^4m)$. Here, we give an $O(n^2m)$-time algorithm for both of these problems. A consequence is a speed up of several known algorithms.

4.2. Large Scale Networks Performance and Modeling

4.2.1. Spatial Interactions of Peers and Performance of File Sharing Systems

Participants: François Baccelli, Fabien Mathieu, Ilkka Norros.

We propose in [24] a new model for peer-to-peer networking which takes the network bottlenecks into account beyond the access. This model allows one to cope with key features of P2P networking like degree or locality constraints or the fact that distant peers often have a smaller rate than nearby peers. We show that the spatial point process describing peers in their steady state then exhibits an interesting repulsion phenomenon. We analyze two asymptotic regimes of the peer-to-peer network: the fluid regime and the hard–core regime. We get closed form expressions for the mean (and in some cases the law) of the peer latency and the download rate obtained by a peer as well as for the spatial density of peers in the steady state of each regime, as well as an accurate approximation that holds for all regimes. The analytical results are based on a mix of mathematical analysis and dimensional analysis and have important design implications. The first of them is the existence of a setting where the equilibrium mean latency is a decreasing function of the load, a phenomenon that we call super-scalability.

4.2.2. User Behavior Modeling: Four Months in DailyMotion

Participants: Yannick Carlinet, The Dang Huynh, Bruno Kauffmann, Fabien Mathieu, Ludovic Noirie, Sébastien Tixeuil.

The growth of User-Generated Content (UGC) traffic makes the understanding of its nature a priority for network operators, content providers and equipment suppliers. In [13], we study a four-month dataset that logs all video requests to DailyMotion made by a fixed subset of users. We were able to infer user sessions from raw data, to propose a Markovian model of these sessions, and to study video popularity and its evolution over time. The presented results are a first step for synthesizing an artificial (but realistic) traffic that could be used in simulations or experimental testbeds.

4.2.3. Multi-Carrier Networks: on the Manipulability of Voting Systems

Participants: François Durand, Fabien Mathieu, Ludovic Noirie.

Today, Internet involves many actors who are making revenues on it (operators, companies, service providers,...). It is therefore important to be able to make fair decisions in this large-scale and highly competitive economical ecosystem. One of the main issues is to prevent actors from manipulating the natural outcome of the decision process. For that purpose, game theory is a natural framework. In that context, voting systems represent an interesting alternative that, to our knowledge, has not yet been considered. They allow competing entities to decide among different options. Strong theoretical results showed that all voting systems are susceptible to be manipulated by one single voter, except for some "degenerated" and non-acceptable cases. However, very little is known about how much a voting system is manipulable in practical scenarios. In [25], we investigate empirically the use of voting systems for choosing end-to-end paths in multi-carrier networks, analyzing their manipulability and their economical efficiency. We show that one particular system, called Single Transferable Vote (STV), is largely more resistant to manipulability than the natural system which tries to get the economical optimum. Moreover, STV manages to select paths close to the economical optimum, whether the participants try to cheat or not.

4.3. Fault Tolerance in Distributed Networks

4.3.1. Wait-Freedom with Advice

Participants: Carole Delporte-Gallet, Hugues Fauconnier, Eli Gafni, Petr Kuznetsov.

In [14], we motivate and propose a new way of thinking about failure detectors which allows us to define, quite surprisingly, what it means to solve a distributed task *wait-freeusing a failure detector*. In our model, the system is composed of *computation* processes that obtain inputs and are supposed to output in a finite number of steps and *synchronization* processes that are subject to failures and can query a failure detector. We assume that, under the condition that *correct* synchronization processes take sufficiently many steps, they provide the computation processes with enough *advice* to solve the given task wait-free: every computation process outputs in a finite number of its own steps, regardless of the behavior of other computation processes. Every task can thus be characterized by the *weakest* failure detector that allows for solving it, and we show that every such failure detector captures a form of set agreement. We then obtain a complete classification of tasks, including ones that evaded comprehensible characterization so far, such as renaming or weak symmetry breaking.

4.3.2. Partial synchrony based on set timeliness

Participants: Markos Aguilera, Carole Delporte-Gallet, Hugues Fauconnier, Sam Toueg.

We introduce in [1], a new model of partial synchrony for read-write shared memory systems. This model is based on the simple notion of set timeliness—a natural generalization of the seminal concept of timeliness in the partially synchrony model of Dwork et al. (J. ACM 35(2):288–323, 1988). Despite its simplicity, the concept of set timeliness is powerful enough to define a family of partially synchronous systems that closely match individual instances of the $t-$resilient $k-$set agreement problem among n processes, henceforth denoted $(t, k, n)-$agreement. In particular, we use it to give a partially synchronous system that is synchronous enough for solving $(t, k, n)-$ agreement, but not enough for solving two incrementally stronger problems, namely, $(t + 1, k, n)-$agreement, which has a slightly stronger resiliency requirement, and $(t, k-1, n)-$agreement, which has a slightly stronger agreement requirement. This is the first partially synchronous system that separates these sub-consensus problems. The above results show that set timeliness can be used to study and compare the partial synchrony requirements of problems that are strictly weaker than consensus.

4.3.3. Byzantine Agreement with Homonyms in Synchronous Systems

Participants: Carole Delporte-Gallet, Hugues Fauconnier, Hung Tran-The.

In [15], [6], we consider the Byzantine agreement problem (BA) in synchronous systems with homonyms. In this model different processes may have the same authenticated identifier. In such a system of n processes sharing a set of l identifiers, we define a distribution of the identifiers as an integer partition of n into l parts $n_1, ..., n_l$ giving for each identifier i the number of processes having this identifier.

Assuming that the processes know the distribution of identifiers we give a necessary and sufficient condition on the integer partition of n to solve the Byzantine agreement with at most t Byzantine processes. Moreover we prove that there exists a distribution of l identifiers enabling to solve Byzantine agreement with at most t Byzantine processes if and only if $n > 3t$, $l > t$ and where $r = n \mod l$.

This bound is to be compared with the $l > 3t$ bound proved in Delporte-Gallet et al. (2011) when the processes do not know the distribution of identifiers.

4.3.4. Homonyms with forgeable identifiers

Participants: Carole Delporte-Gallet, Hugues Fauconnier, Hung Tran-The.

In [16], we refine the Byzantine Agreement problem (BA) in synchronous systems with homonyms, in the particular case where some identifiers may be forgeable. More precisely, the n processes share a set of l $(1 \le l \le n)$ identifiers. Assuming that at most t processes may be Byzantine and at most k $(t \le k \le l)$ of these identifiers are forgeable in the sense that any Byzantine process can falsely use them, we prove that Byzantine Agreement problem is solvable if and only if $l > 2t + k$. Moreover we extend this result to systems with authentication by signatures in which at most k signatures are forgeable and we prove that Byzantine Agreement problem is solvable if and only if $l > t + k$.

4.4. Discrete Optimization Algorithms

4.4.1. Estimating satisfiability

Participants: Yacine Boufkhad, Thomas Hugel.

The problem of estimating the proportion of satisfiable instances of a given CSP (constraint satisfaction problem) can be tackled through weighting. It consists in putting onto each solution a non-negative real value based on its neighborhood in a way that the total weight is at least 1 for each satisfiable instance. We define in [3], a general weighting scheme for the estimation of satisfiability of general CSPs. First we give some sufficient conditions for a weighting system to be correct. Then we show that this scheme allows for an improvement on the upper bound on the existence of non-trivial cores in 3-SAT obtained by Maneva and Sinclair (2008) to 4.419. Another more common way of estimating satisfiability is ordering. This consists in putting a total order on the domain, which induces an orientation between neighboring solutions in a way that prevents circuits from appearing, and then counting only minimal elements. We compare ordering and weighting under various conditions.

4.4.2. Attractive force search algorithm for piecewise convex maximization problems

Participants: Dominique Fortin, Ider Tseveendorj.

In [8], we consider mathematical programming problems with the so-called piecewise convex objective functions. A solution method for this interesting and important class of nonconvex problems is presented. This method is based on Newton's law of universal gravitation, multicriteria optimization and Helly's theorem on convex bodies. Numerical experiments using well known classes of test problems on piecewise convex maximization, convex maximization as well as the maximum clique problem show the efficiency of the approach.

4.4.3. B-spline interpolation: Toeplitz inverse under corner perturbations

Participant: Dominique Fortin.

For Toeplitz matrices associated with degree 3 and 4 uniform B-spline interpolation, the inverse may be analytically known [7], saving the standard inverse calculations. It generalizes to any degree as a row of the Eulerian numbers triangle.

GRAND-LARGE Project-Team

5. New Results

5.1. Communication avoiding algorithms for linear algebra

Participants: Laura Grigori, Amal Khabou, Mathias Jacquelin, Sophie Moufawad.

The focus of this research is on the design of efficient parallel algorithms for solving problems in numerical linear algebra, as solving very large sets of linear equations and large least squares problems, often with millions of rows and columns. These problems arise in many numerical simulations, and solving them is very time consuming.

Our research focuses on developing new algorithms for linear algebra problems, that minimize the required communication, in terms of both latency and bandwidth. We have introduced in 2008 two communication avoiding algorithms for computing the LU and QR factorizations, that we refer to as CALU and CAQR (joint work with J. Demmel and M. Hoemmen from U.C. Berkeley, J. Langou from C.U. Denver, and H. Xiang then at Inria) [18] [8]. Since then, we continue designing communication avoiding algorithm for other operations in both dense and sparse linear algebra. The communication avoiding algorithms are now studied by several other groups, including groups at Inria, and they start being implemented and being available in public libraries as ScaLAPACK.

During 2012, our research [43] has focused on the design of the LU decomposition with panel rank revealing pivoting (LU_PRRP), an LU factorization algorithm based on strong rank revealing QR panel factorization. LU_PRRP is more stable than Gaussian elimination with partial pivoting (GEPP), with a theoretical upper bound of the growth factor of $(1+\tau b)^{(n/b)-1}$, where b is the size of the panel used during the block factorization, τ is a parameter of the strong rank revealing QR factorization, and n is the number of columns of the matrix. For example, if the size of the panel is $b = 64$, and $\tau = 2$, then $(1+2b)^{(n/b)-1} = (1.079)^{n-1} \ll 2^{n-1}$, where 2^{n-1} is the upper bound of the growth factor of GEPP. Our extensive numerical experiments show that the new factorization scheme is as numerically stable as GEPP in practice, but it is more resistant to pathological cases. The LU_PRRP factorization does only $O(n^2 b)$ additional floating point operations compared to GEPP. We have also introduced CALU_PRRP, a communication avoiding version of LU_PRRP that minimizes communication. CALU_PRRP is based on tournament pivoting, with the selection of the pivots at each step of the tournament being performed via strong rank revealing QR factorization. CALU_PRRP is more stable than CALU, the communication avoiding version of GEPP, with a theoretical upper bound of the growth factor of $(1+\tau b)^{\frac{n}{b}(H+1)-1}$, where H is the height of the reduction tree used during tournament pivoting. The upper bound of the growth factor of CALU is $2^{n(H+1)-1}$. CALU_PRRP is also more stable in practice and is resistant to pathological cases on which GEPP and CALU fail.

Our work has also focused on designing algorithms that are optimal over multiple levels of memory hierarchy and parallelism. In [32] we present an algorithm for performing the LU factorization of dense matrices that is suitable for computer systems with two levels of parallelism. This algorithm is able to minimize both the volume of communication and the number of messages transferred at every level of the two-level hierarchy of parallelism. We present its implementation for a cluster of multicore processors based on MPI and Pthreads. We show that this implementation leads to a better performance than routines implementing the LU factorization in well-known numerical libraries. For matrices that are tall and skinny, that is they have many more rows than columns, our algorithm outperforms the corresponding algorithm from ScaLAPACK by a factor of 4.5 on a cluster of 32 nodes, each node having two quad-core Intel Xeon EMT64 processors.

5.2. Preconditioning techniques for solving large systems of equations

Participants: Laura Grigori, Riadh Fezzanni, Sophie Moufawad.

A different direction of research is related to preconditioning large sparse linear systems of equations. This research is performed in the context of ANR PETALh project (2011-2012), which follows the ANR PETAL project (2008-2009). It is conducted in collaboration with Frederic Nataf from University Paris 6.

Several highly used preconditioners are for example the incomplete LU factorizations and Schwarz based approaches as used in domain decomposition. Most of these preconditioners are known to have scalability problems. The number of iterations can increase significantly when the size of the problem increases or when the number of independent domains is increased. This is often due to the presence of several low frequency modes that hinder the convergence of the iterative method. To address this problem, we study a different class of preconditioners, called direction preserving or filtering preconditioners. These preconditioners have the property of being identical to the input matrix on a given filtering vector. A judicious choice of the vector allows to alleviate the effect of low frequency modes on the convergence.

We consider in particular two classes of preconditioners. The first preconditoner is an incomplete decomposition that satisfies the filtering property [13]. The nested preconditioner has the same property for a specific vector of all ones. However the construction is different and takes advantage of a nested structure of the input matrix. The previous research on these methods considered only matrices arising from the discretization of PDEs on structured grids, where the matrix has a block tridiagonal structure. This structure imposes a sequential computation of the preconditioner and it is not suitable for the more general case of unsructured grids. Hence, while very efficient, the usage of these preconditioners was very limited. At the beginning of this research we have obtained several theoretical results for these methods that demonstrate their numerical behavior and convergence properties for cases arising from the discretization of PDEs on structured grids [13]. But the main result is the development of a generalized method [10], [11] that has two important properties: it allows the filtering property to be satisfied for any input matrix; the matrix can be reordered such that its computation is highly parallel. Experimental results show that the method is very efficient for certain classes of matrices, and shows good scalability results in terms of both problem size and number of processors. In addition to finalizing this work, our research also focused on extending the block filtering factorization to include other approximation techniques that allowed us to introduce a parameter whose tuning permits to solve very difficult problems.

5.3. MIcrowave Data Analysis for petaScale computers

Participants: Laura Grigori, Mikolaj Szydlarski, Meisam Shariffy.

Generalized least square problems with non-diagonal weights arise frequently in an estimation of two dimensional images from data of cosmological as well as astro- or geo- physical observations. As the observational data sets keep growing at Moore's rate, with their volumes exceeding tens and hundreds billions of samples, the need for fast and efficiently parallelizable iterative solvers is generally recognized.

In this work [36] we propose a new iterative algorithm for solving generalized least square systems with weights given by a block-diagonal matrix with Toeplitz blocks. Such cases are physically well motivated and correspond to measurement noise being piece-wise stationary – a common occurrence in many actual observations. Our iterative algorithm is based on the conjugate gradient method and includes a parallel two-level preconditioner (2lvl-PCG) constructed from a limited number of sparse vectors estimated from the coefficients of the initial linear system.

Our prototypical application is the map-making problem in the Cosmic Microwave Background data analysis. We show experimentally that our parallel implementation of 2lvl-PCG outperforms by a factor of up to 6 the standard one-level PCG in terms of both the convergence rate and the time to solution on up to $12,228$ cores of NERSC's Cray XE6 (Hopper) system displaying nearly perfect strong and weak scaling behavior in this regime.

5.4. Innovative linear system solvers for hybrid multicore/GPU architectures

Participant: Marc Baboulin.

The advent of new processor architectures (e.g. multicore, GPUs) requires the rethinking of most of the scientific applications and innovative methods must be proposed in order to take full advantage of current supercomputers [14].

To accelerate linear algebra solvers on current parallel machines, we introduced in public domain libraries a class of solvers based on statistical techniques. A first application concerns the solution of a square linear systems $Ax = b$. We study a random transformation of A that enables us to avoid pivoting and then to reduce the amount of communication [16]. Numerical experiments show that this randomization can be performed at a very affordable computational price while providing us with a satisfying accuracy when compared to partial pivoting. This random transformation called Partial Random Butterfly Transformation (PRBT) is optimized in terms of data storage and flops count. In the solver that we developed, PRBT combined with LU factorization with no pivoting take advantage of the latest generation of hybrid multicore/GPU machines and outperform existing factorization routines from current parallel library MAGMA.

A second application is related to solving symmetric indefinite systems via LDL^T factorization for which there was no existing parallel implementation in the dense library ScaLAPACK. We developed an efficient and innovative parallel tiled algorithm for solving symmetric indefinite systems on multicore architectures [54]& [25]. This solver avoids pivoting by using a multiplicative preconditioning based on symmetric randomization. This randomization prevents the communication overhead due to pivoting, is computationally inexpensive and requires very little storage. Following randomization, a tiled LDLT factorization is used that reduces synchronization by using static or dynamic scheduling. We compare Gflop/s performance of our solver with other types of factorizations on a current multicore machine and we provide tests on accuracy using LAPACK test cases.

5.5. MILEPOST GCC: machine learning enabled self-tuning compiler

Participant: Grigori Fursin [correspondant].

Tuning compiler optimizations for rapidly evolving hardware makes porting and extending an optimizing compiler for each new platform extremely challenging. Iterative optimization is a popular approach to adapting programs to a new architecture automatically using feedback-directed compilation. However, the large number of evaluations required for each program has prevented iterative compilation from widespread take-up in production compilers. Machine learning has been proposed to tune optimizations across programs systematically but is currently limited to a few transformations, long training phases and critically lacks publicly released, stable tools.

Our approach is to develop a modular, extensible, self-tuning optimization infrastructure to automatically learn the best optimizations across multiple programs and architectures based on the correlation between program features, run-time behavior and optimizations. In this paper we describe MILEPOST GCC, the first publicly-available open-source machine learning-based compiler. It consists of an Interactive Compilation Interface (ICI) and plugins to extract program features and exchange optimization data with the cTuning.org open public repository. It automatically adapts the internal optimization heuristic at function-level granularity to improve execution time, code size and compilation time of a new program on a given architecture. Part of the MILEPOST technology together with low-level ICI-inspired plugin framework is now included in the mainline GCC.

We developed machine learning plugins based on probabilistic and transductive approaches to predict good combinations of optimizations. Our preliminary experimental results show that it is possible to automatically reduce the execution time of individual MiBench programs on various machines from GRID5000, some by more than a factor of 2, while also improving compilation time and code size. We also present a realistic multi-objective optimization scenario for Berkeley DB library using MILEPOST GCC and improve execution time by approximately 17%, while reducing compilation time and code size by 12% and 7% respectively on Intel Xeon processor.

5.6. Loop Transformations: Convexity, Pruning and Optimization

Participant: Cédric Bastoul.

High-level loop transformations are a key instrument in mapping computational kernels to effectively exploit resources in modern processor architectures. However, determining appropriate compositions of loop transformations to achieve this remains a significantly challenging task; current compilers may achieve significantly lower performance than hand-optimized programs. To address this fundamental challenge, we first present a convex characterization of all distinct, semantics-preserving, multidimensional affine transformations. We then bring together algebraic, algorithmic, and performance analysis results to design a tractable optimization algorithm over this highly expressive space. The framework has been implemented and validated experimentally on a representative set of benchmarks run on state-of-the-art multi-core platforms.

5.7. Non-self-stabilizing and self-stabilizing gathering in networks of mobile agents–the notion of speed

Participants: Joffroy Beauquier, Janna Burman, Julien Clment, Shay Kutten.

In the population protocol model, each agent is represented by a finite state machine. Agents are anonymous and supposed to move in an asynchronous way. When two agents come into range of each other (“meet”), they can exchange information. One of the vast variety of motivating examples to the population protocols model is ZebraNet. ZebraNet is a habitat monitoring application where sensors are attached to zebras and collect biometric data (e.g. heart rate, body temperature) and information about their behavior and migration patterns (via GPS). The population protocol model is, in some sense, related to cloud computing and to networks characterized by asynchrony,large scale, the possibility of failures, in the agents as well as in the communications, with the constraint that each agent is resource limited.

In order to extend the computation power and efficiency of the population protocol model, various extensions were suggested. Our contribution is an extension of the population protocol model that introduces the notion of “speed”, in order to capture the fact that the mobile agents move at different speeds and/or have different communication ranges and/or move according to different patterns and/or visit different places with different frequencies. Intuitively, fast agents which carry sensors with big communication ranges communicate with other agents more frequently than other agents do. This notion is formalized by allocating a cover time, cv, to each mobile agent v. cv is the minimum number of events in the whole system that occur before agent v meets every other agent at least once. As a fundamental example, we have considered the basic problem of gathering information that is distributed among anonymous mobile agents and where the number of agents is unknown. Each mobile agent owns a sensed input value and the goal is to communicate the values (as a multi-set, one value per mobile agent) to a fixed non-mobile base station (BS), with no duplicates or losses.

Gathering is a building block for many monitoring applications in networks of mobile agents. For example, a solution to this problem can solve a transaction commit/abort task in MANETs, if the input values of agents are votes (and the number of agents is known to BS). Moreover, the gathering problem can be viewed as a formulation of the routing problem in Disruption Tolerant Networks.

We gave different solutions to the gathering in the model of mobile agents with speed and we proved that one of them is optimal.

5.8. Making Population Protocols Self-stabilizing

Participants: Joffroy Beauquier, Janna Burman, Shay Kutten, Brigitte Rozoy.

As stated in the previous paragraph, the application domains of the population protocol model are asynchronous large scale networks, in which failures are possible and must be taken into account. This work concerns failures and namely the technique of self-stabilization for tolerating them.

Developing self-stabilizing solutions (and proving them) is considered to be more challenging and complicated than developing classical solutions, where a proper initialization of the variables can be assumed. This remark holds for a large variety of models and hence, to ease the task of the developers, some automatic techniques have been proposed to transform programs into self-stabilizing ones.

We have proposed such a transformer for algorithms in the population protocol model introduced for dealing with resource-limited mobile agents. The model we consider is a variation of the original one in that there is a non mobile agent, the base station, and that the communication characteristics (e.g. moving speed, communication radius) of the agents are considered through the notion of cover time.

The automatic transformer takes as an input an algorithm solving a static problem and outputs a self-stabilizing solution for the same problem. To the best of our knowledge, it is the first time that such a transformer for self-stabilization is presented in the framework of population protocols. We prove that the transformer we propose is correct and we make the complexity analysis of the stabilization time.

5.9. Self-stabilizing synchronization in population protocols with cover times

Participants: Joffroy Beauquier, Janna Burman, Shay Kutten, Brigitte Rozoy.

Synchronization is widely considered as an important service in distributed systems which may simplify protocol design. Phase clock is a general synchronization tool that provides a form of a logical time. We have developed a self-stabilizing phase clock algorithm suited to the model of population protocols with cover time. We have shown that a phase clock is impossible in the model with only constant-state agents. Hence, we assumed an existence of resource unlimited agent - the base station. The clock size and duration of each phase of the proposed phase clock tool are adjustable by the user. We provided application examples of this tool and demonstrate how it can simplify the design of protocols. In particular, it yields a solution to Group Mutual Exclusion problem.

5.10. Impossibility of consensus for population protocol with cover times

Participants: Joffroy Beauquier, Janna Burman.

We have extended the impossibility result for asynchronous consensus of Fischer, Lynch and Paterson (FLP) to the asynchronous model of population protocols with cover times. We noted that the proof of FLP does not apply. Indeed, the key lemma stating that two successive factors in an execution, involving disjoint subsets of agents, commute, is no longer true, because of the cover time property. Then we developed a completely different approach and we proved that there is no general solution to consensus for population protocols with cover times, even if there is a single possible crash. We noted that this impossibility result also applies to randomized asynchronous consensus, contrary to what happens in the classical message-passing or shared memory communication models, in which the problem is solvable inside some bounds on the number of faulty processes. Then, for circumventing these impossibility results, we introduced the phase clock oracle and the S oracle, and we shown how they allow to design solutions.

5.11. Routing and synchronization in large scale networks of very cheap mobile sensors

Participants: Joffroy Beauquier, Brigitte Rozoy.

In a next future, large networks of very cheap mobile sensors will be deployed for various applications, going from wild life preserving or environmental monitoring up to medical or industrial system control. Each sensor will cost only a few euros, allowing a large scale deployment. They will have only a few bit of memory, no identifier, weak capacities of computation and communication, no real time clock and will be prone to failures. Moreover such networks will be fundamentally dynamic. The goal of this subject is to develop the basic protocols and algorithms for rudimentary distributed systems for such networks. The studied problems are basic ones, like data collection, synchronization (phase clock, mutual exclusion, group mutual exclusion), fault tolerance (consensus), automatic transformers, always in a context of possible failures. A well known model has already been proposed for such networks, the population protocol model. In this model, each sensor is represented by a finite state machine. Sensors are anonymous and move in an asynchronous way. When two sensors come into range of each other ("meet"), they can exchange information. One of the vast variety of motivating examples for this model is ZebraNet. ZebraNet is a habitat monitoring application in which sensors are attached to zebras in order to collect biometric data (e.g., heart rate, body temperature) and information about their behavior and migration patterns. Each pair of zebras meets from time to time. During such meetings (events), ZebraNet's agents (zebras' attached sensors) exchange data. Each agent stores its own sensor data as well as data of other sensors that were in range in the past. They upload data to a base station whenever it is nearby. It was shown that the set of applications that can be solved in the original model of population protocols is rather limited. Other models (such as some models of Delay/Disruption-Tolerant Networks - DTNs), where each node maintains links and connections even to nodes it may interact with only intermittently, do not seem to suit networks with small memory agents and a very large (and unknown) set of anonymous agents. That is why we enhance the model of population protocols by introducing a notion of "speed". We try to capture the fact that the mobile agents move at different speeds and/or have different communication ranges and/or move according to different patterns and/or visit different places with different frequencies. Intuitively, fast agents which carry sensors with large communication ranges communicate with other agents more frequently than other agents do. This notion is formalized by the notion of cover time for each agent. The cover time of an agent is the unknown number of events (pairwise meetings) in the whole system that occur (during any execution interval) before agent v meets every other agent at least once. The model we propose is somehow validated by some recent statistical results, obtained from empirical data sets regarding human or animal mobility. An important consequence of our approach is that the analytic complexity of the protocols designed in this model is possible, independently of any simulation or experimentation. For instance, we consider the fundamental problem of gathering different pieces of information, each sensed by a different anonymous mobile agent, and where the number of agents is unknown. The goal is to communicate the sensed values (as a multi-set, one value per mobile agent) to a base station, with no duplicates or losses. Gathering is a building block for many monitoring applications in networks of mobile agents. Moreover, the gathering problem can be viewed as a special case of the routing problem in DTNs, in which there is only one destination, the base station. Then we are able to compute the complexity of solutions we propose, as well as those of solutions used in experimental projects (like ZebraNet), and to compare them. The algorithms we present are self-stabilizing. Such algorithms have the important property of operating correctly regardless of their initial state (except for some bounded period). In practice, self-stabilizing algorithms adjust themselves automatically to any changes or corruptions of the network components (excluding the algorithm's code). These changes are assumed to cease for some sufficiently long period. Self-stabilization is considered for two reasons. First, mobile agents are generally fragile, subject to failures and hard to initialize. Second, systems of mobile agents are by essence dynamic, some agents leave the system while new ones are introduced. Self-stabilization is a well adapted framework for dealing with such situations.

5.12. Self-Stabilizing Control Infrastructure for HPC

Participants: Thomas Hérault, Camille Coti.

High performance computing platforms are becoming larger, leading to scalability and fault-tolerance issues for both applications and runtime environments (RTE) dedicated to run on such machines. After being deployed, usually following a spanning tree, a RTE needs to build its own communication infrastructure to manage and monitor the tasks of parallel applications. Previous works have demonstrated that the Binomial Graph topology (BMG) is a good candidate as a communication infrastructure for supporting scalable and fault-tolerant RTE.

In this work, we presented and analyzed a self-stabilizing algorithm to transform the underlying communication infrastructure provided by the launching service (usually a tree, due to its scalability during launch time) into a BMG, and maintain it in spite of failures. We demonstrated that this algorithm is scalable, tolerates transient failures, and adapts itself to topology changes.

The algorithms are scalable, in the sense that all process memory, number of established communication links, and size of messages are logarithmic with the number of elements in the system. The number of synchronous rounds to build the system is also logarithmic, and the number of asynchronous rounds in the worst case is square logarithmic with the number of elements in the system. Moreover, the salf-stabilizing property of the algorithms presented induce fault-tolerance and self-adaptivity. Performance evaluation based on simulations predicts a fast convergence time (1/33s for 64K nodes), exhibiting the promising properties of such self-stabilizing approach.

We pursue this work by implementing and evaluating the algorithms in the STCI runtime environment to validate the theoretical results.

5.13. Large Scale Peer to Peer Performance Evaluations

Participant: Serge Petiton.

5.13.1. Large Scale Grid Computing

Recent progress has made possible to construct high performance distributed computing environments, such as computational grids and cluster of clusters, which provide access to large scale heterogeneous computational resources. Exploration of novel algorithms and evaluation of performance is a strategic research for the future of computational grid scientific computing for many important applications [82]. We adapted [63] an explicit restarted Lanczos algorithm on a world-wide heterogeneous grid platform. This method computes one or few eigenpairs of a large sparse real symmetric matrix. We take the specificities of computational resources into account and deal with communications over the Internet by means of techniques such as out-of-core and data persistence. We also show that a restarted algorithm and the combination of several paradigms of parallelism are interesting in this context. We perform many experimentations using several parameters related to the Lanczos method and the configuration of the platform. Depending on the number of computed Ritz eigenpairs, the results underline how critical the choice of the dimension of the working subspace is. Moreover, the size of platform has to be scaled to the order of the eigenproblem because of communications over the Internet.

5.13.2. High Performance Cluster Computing

Grid computing focuses on making use of a very large amount of resources from a large-scale computing environment. It intends to deliver high-performance computing over distributed platforms for computation and data-intensive applications. We propose [93] an effective parallel hybrid asynchronous method to solve large sparse linear systems by the use of a Grid Computing platform Grid5000. This hybrid method combines a parallel GMRES(m) (Generalized Minimum RESidual) algorithm with the Least Square method that needs some eigenvalues obtained from a parallel Arnoldi algorithm. All of these algorithms run on the different processors of the platform Grid5000. Grid5000, a 5000 CPUs nation-wide infrastructure for research in Grid computing, is designed to provide a scientific tool for computing. We discuss the performances of this hybrid method deployed on Grid5000, and compare these performances with those on the IBM SP series supercomputers.

5.13.3. Large Scale Power aware Computing

Energy conservation is a dynamic topic of research in High Performance Computing and Cluster Computing. Power-aware computing for heterogeneous world-wide Grid is a new track of research. We have studied and evaluated the impact of the heterogeneity of the computing nodes of a Grid platform on the energy consumption. We propose to take advantage of the slack-time caused by the heterogeneity in order to save energy with no significant loss of performance by using Dynamic Voltage Scaling (DVS) in a distributed eigensolver [64]. We show that using DVS only during the slack-time does not penalize the performances but it does not provide significant energy savings. If DVS is applied to all the execution, we get important global and local energy savings (respectively up to 9% and 20%) without a significant rise of the wall-clock times.

5.14. High Performance Linear Algebra on the Grid

Participants: Thomas Hérault, Camille Coti.

Previous studies have reported that common dense linear algebra operations do not achieve speed up by using multiple geographical sites of a computational grid. Because such operations are the building blocks of most scientific applications, conventional supercomputers are still strongly predominant in high-performance computing and the use of grids for speeding up large-scale scientific problems is limited to applications exhibiting parallelism at a higher level.

In this work, we have identified two performance bottlenecks in the distributed memory algorithms implemented in ScaLAPACK, a state-of-the-art dense linear algebra library. First, because ScaLAPACK assumes a homogeneous communication network, the implementations of ScaLAPACK algorithms lack locality in their communication pattern. Second, the number of messages sent in the ScaLAPACK algorithms is significantly greater than other algorithms that trade flops for communication.

This year, we presented a new approach for computing a QR factorization one of the main dense linear algebra kernels of tall and skinny matrices in a grid computing environment that overcomes these two bottlenecks. Our contribution is to articulate a recently proposed algorithm (Communication Avoiding QR) with a topology-aware middleware (QCG-OMPI) in order to confine intensive communications (ScaLAPACK calls) within the different geographical sites.

An experimental study conducted on the Grid5000 platform shows that the resulting performance increases linearly with the number of geographical sites on large-scale problems (and is in particular consistently higher than ScaLAPACKs).

5.15. Emulation of Volatile Systems

Participants: Thomas Largillier, Benjamin Quetier, Sylvain Peyronnet, Thomas Hérault, Franck Cappello.

In the process of developping grid applications, people need to often evaluate the robustness of their work. Two common approaches are simulation, where one can evaluate his software and predict behaviors under conditions usually unachievable in a laboratory experiment, and experimentation, where the actual application is launched on an actual grid. However simulation could ignore unpredictable behaviors due to the abstraction done and experimation does not guarantee a controlled and reproducible environment, and simulation often introduces a high level of abstraction that make the discovery and study of unexpected, but real, behaviors a rare event.

In this work, we proposed an emulation platform for parallel and distributed systems including grids where both the machines and the network are virtualized at a low level. The use of virtual machines allows us to test highly accurate failure injection since we can destroy virtual machines, and network virtualization provides low-level network emulation. Failure accuracy is a criteria that evaluates how realistic a fault is. The accuracy of our framework has been evaluated through a set of micro benchmarks and a very stable P2P system called Pastry.

We are in the process of developping a fault injection tool to work with the platform. it will be an extension of the work started in the tool Fail. The interest of this work is that using Xen virtual machines will allow to model strong adversaries since it is possible to have virtual machines with shared memory. These adversaries will be stronger since they will be able to use global fault injection strategies.

5.16. Exascale Systems

Participant: Franck Cappello.

Over the last 20 years, the open-source community has provided more and more software on which the world's high-performance computing systems depend for performance and productivity. The community has invested millions of dollars and years of effort to build key components. Although the investments in these separate software elements have been tremendously valuable, a great deal of productivity has also been lost because of the lack of planning, coordination, and key integration of technologies necessary to make them work together smoothly and efficiently, both within individual petascale systems and between different systems. A repository gatekeeper and an email discussion list can coordinate open-source development within a single project, but there is no global mechanism working across the community to identify critical holes in the overall software environment, spot opportunities for beneficial integration, or specify requirements for more careful coordination. It seems clear that this completely uncoordinated development model will not provide the software needed to support the unprecedented parallelism required for peta/exascale computation on millions of cores, or the flexibility required to exploit new hardware models and features, such as transactional memory, speculative execution, and GPUs. We presented a rational promoting that the community must work together to prepare for the challenges of exascale computing, ultimately combing their efforts in a coordinated International Exascale Software Project.

Over the past few years resilience has became a major issue for high-performance computing (HPC) systems, in particular in the perspective of large petascale systems and future exascale systems. These systems will typically gather from half a million to several millions of central processing unit (CPU) cores running up to a billion threads. From the current knowledge and observations of existing large systems, it is anticipated that exascale systems will experience various kind of faults many times per day. It is also anticipated that the current approach for resilience, which relies on automatic or application level checkpoint/restart, will not work because the time for checkpointing and restarting will exceed the mean time to failure of a full system. This set of projections leaves the community of fault tolerance for HPC systems with a difficult challenge: finding new approaches, which are possibly radically disruptive, to run applications until their normal termination, despite the essentially unstable nature of exascale systems. Yet, the community has only five to six years to solve the problem. In order to start addressing this challenge, we synthesized the motivations, observations and research issues considered as determinant of several complimentary experts of HPC in applications, programming models, distributed systems and system management.

As a first step to adress the resilience challenge, we conducted a comprehensive study of the state of the art . The emergence of petascale systems and the promise of future exascale systems have reinvigorated the community interest in how to manage failures in such systems and ensure that large applications, lasting several hours or tens of hours, are completed successfully. Most of the existing results for several key mechanisms associated with fault tolerance in high-performance computing (HPC) platforms follow the rollback-recovery approach. Over the last decade, these mechanisms have received a lot of attention from the community with different levels of success. Unfortunately, despite their high degree of optimization, existing approaches do not fit well with the challenging evolutions of large-scale systems. There is room and even a need for new approaches. Opportunities may come from different origins: diskless checkpointing, algorithmic-based fault tolerance, proactive operation, speculative execution, software transactional memory, forward recovery, etc. We provided the following contributions: (1) we summarize and analyze the existing results concerning the failures in large-scale computers and point out the urgent need for drastic improvements or disruptive approaches for fault tolerance in these systems; (2) we sketch most of the known opportunities and analyze their associated limitations; (3) we extract and express the challenges that the HPC community will have to face for addressing the stringent issue of failures in HPC systems.

HIEPACS Project-Team

6. New Results

6.1. Algorithms and high-performance solvers

6.1.1. Dense linear algebra solvers for multicore processors accelerated with multiple GPUs

In collaboration with the Inria RUNTIME team and the University of Tennessee, we have designed dense linear algebra solvers that can fully exploit a node composed of a multicore processor accelerated with multiple GPUs. This work has been integrated in the latest release of the MAGMA package (http://icl.cs.utk.edu/magma/). We have used the StarPU runtime system to ensure the portability of our algorithms and codes. We have also investigated the case of partial pivoting LU factorization. The pivot selection induces a large number of low granularity tasks which are a potential bottleneck when handled with a runtime system; we have thus designed methods which aim at limiting the number of tasks.

6.1.2. Task-based Conjugate-Gradient for multi-GPUs platforms

Whereas most today parallel High Performance Computing (HPC) software is written as highly tuned code taking care of low-level details, the advent of the manycore area forces the community to consider modular programming paradigms and delegate part of the work to a third party software. That latter approach has been shown to be very productive and efficient with regular algorithms, such as dense linear algebra solvers. In this paper we show that such a model can be efficiently applied to a much more irregular and less compute intensive algorithm. We illustrate our discussion with the standard unpreconditioned Conjugate Gradient (CG) that we carefully express as a task-based algorithm. We use the StarPU runtime system to assess the efficiency of the approach on a computational platform consisting of three NVIDIA Fermi GPUs. We show that almost optimum speed up (up to 2.89) may be reached (relatively to a mono-GPU execution) when processing large matrices and that the performance is portable when changing the low-level memory transfer mechanism. This work is developed in the framework of the PhD of Stojce Nakov.

6.1.3. Resilience in numerical simulations

Various interpolations strategies to handle restarting Krylov subspace methods in case of core faults have been investigated. The underlying idea is to recover fault entries of the iterate via interpolation from existing values available on neighbor cores. In particular, we design a scheme that enables to preserve the key property of GMRES that is the residual norm monotonicity of the iterates even when failures occur. This work is developed in the framework of Mawussi Zounon's PhD funded by the ANR-RESCUE. Notice that theses activities are also part of our contribution to the G8-ECS (Enabling Climate Simulation at extreme scale).

6.1.4. Block GMRES method with inexact breakdowns and deflated restarting

We have considered the solution of large linear systems with multiple right-hand sides using a block GMRES approach. We designed a new algorithm that effectively handles the situation of almost rank deficient block generated by the block Arnoldi procedure and that enables the recycling of spectral information at restart. The first feature is inherited from an algorithm introduced by Robbé and Sadkane [M. Robbé and M. Sadkane. Exact and inexact breakdowns in the block gmres method. Linear Algebra and its Applications, 419: 265-285, 2006.], while the second one is obtained by extending the deflated restarting strategy proposed by Morgan [R. B. Morgan. Restarted block GMRES with deflation of eigenvalues. Applied Numerical Mathematics, 54(2): 222-236, 2005.]. Through numerical experiments, we have shown that the new algorithm combines the attractive numerical features of its two parents that it outperforms. This work was developed in the framework of the post-doc position of Yan-Fei Jing.

6.1.5. Scalable numerical schemes for scientific applications

For the solution of the elastodynamic equation on meshes with local refinments, we are currently collaborating with Total to design a parallel implementation of a local time refinement technique on top of a discontinuous Galerkin space discretization. This latter technique enables to manage non-conforming meshes suited to deal with multiblock approaches that capture the locally refined regions. this work is developed in the framework of Yohann Dudouit PhD thesis. Perfectly Matched Layers has been designed to cope with the designed numerical scheme and a software prototype for 2D simulation has been implemented.

The calculation of acoustic modes in combustion chambers is a challenging calculation for large 3D geometries. It requires the parallel calculation of a few of the smallest eigenpairs of large unsymmetric matrices in a nonlinear iterative scheme. Various numerical techniques have been considered to attempt recycling spectral information from one nonlinear step to the next that includes Jacobi-Davidson, Krylov-Schur and block Krylov-Schur algorithms. This is part of the PhD research activity of Pablo Salas.

6.1.6. Fast Multipole Methods

Concerning the Fast Multipole Method, our prototype called ScalFMM was completely rewritten in order to easily add new features. There is two main parts: the management of the octree and the parallelization of the method and kernels. This new architecture allow us to easily add new FMM algorithm or kernels and new paradigm of parallelization. The limitation of the classical FMM was that we need all operators (P2M, M2M, M2L, L2L, L2P) on the multipole expansions if we want to add a new kernel. To overcome this and in the context of associated team FastLA, we introduced the black-box FMM algorithm that allow us to be now kernel independent.

6.1.6.1. Optimizations for the M2L operator of the Chebyshev Fast Multipole Method

Most Fast Multipole Methods (FMM) have been developed and optimized for specific kernel functions. Our goal is to improve the efficiency of an FMM that is kernel function independent. The formulation is based on a Chebyshev interpolation scheme and has been studied for asymptotically smooth kernel functions G(x,y) and also for oscillatory ones, such as K(x,y) = G(x,y) exp(ik|x-y|). Two weak points of this formulation are the expensive precomputation of the M2L operators and the higher computational intensity compared to other FMMs. We focused our recent research on these issues. We have come up with a set of optimizations that exploit symmetries far-field interactions and blocking schemes that pave the road for highly optimized matrix-matrix product implementations. Recall, the scope of the FMM as an algorithm to perform fast matrix-vector products (Ax = y) may be twofold: on one hand the result (y) and on the other hand the solution (x). A fast precomputation is crucial in the first and fast running times in the second case. We proposed optimizations that provide more than 1000 times faster precomputation, much less memory requirement and much faster running times than before. All these results are submitted in Journal of computational Physics [27].

6.1.6.2. Pipelining the Chebyshev Fast Multipole Method over a runtime system

Fast Multipole Method are a fundamental operation for the simulation of many physical problems. The high performance design of such methods usually requires to carefully tune the algorithm for both the targeted physics and the hardware. For the Chebyshev Fast Multipole Method (black-box FMM) we have proposed a new approach that achieves high performance across heterogeneous architectures. Our method consists of expressing the Fast Multipole Method algorithm as a task flow and employing a state-of-the-art runtime system, StarPU, in order to process the tasks on the different processing units. We carefully design the task flow, the mathematical operators, their Central Processing Unit (CPU) and Graphics Processing Unit (GPU) implementations, as well as scheduling schemes. We compute potentials and forces of 200 million particles in 48.7 seconds on a homogeneous 160 cores SGI Altix UV 100 and of 30 million particles in 10.9 seconds on a heterogeneous 12 cores Intel Nehalem processor enhanced with 3 Nvidia M2090 Fermi GPUs. All these results are available in [24].

6.2. Efficient algorithmics for code coupling in complex simulations

Dynamic load balancing is an important step conditioning the performance of parallel adaptive codes whose load evolution is difficult to predict. Most of the studies which answer this problem perform well, but are limited to an initially fixed number of processors which is not modified at runtime. These approaches can be very inefficient, especially in terms of resource consumption, as demonstrated by Iqbal et al. As computation progresses, the global workload may increase drastically, exceeding memory limit for instance. In such a case, we argue it should be relevant to adjust the number of processors while maintaining the load balanced. However, this is still an open question that we currently focus on.

To overcome this issue, we propose a new graph repartitioning algorithm, which accepts a variable number of processors, assuming the load is already balanced. We call this problem the $M \times N$ graph repartitioning problem, with M the number of former parts and N the number of newer parts. Our algorithm minimizes both data communication (i.e., cut size) and data migration overheads, while maintaining the computational load balance in parallel. This algorithm is based on a theoretical result, that constructs optimal communication matrices with both a minimum migration volume and a minimum number of communications. It uses recent graph/hypergraph partitioning techniques with fixed vertices in a similar way than the one used in Zoltan for dynamic load-balancing of adaptive simulations. We validate this work for a large variety of real-life graphs (i.e., university of Florida sparse matrix collection), comparing it against state-of-the-art partitioners (Metis, Scotch, Zoltan).

We are considering several perspectives to our work. First, we focus on graph repartitioning in the more general case where both the load and the number of processors vary. We expect this work to be really suitable for next generation of adaptive codes. Finally, to be useful in real-life applications, our algorithm needs to work in parallel, that mainly requires to use a direct *k-way* parallel partitioning software that handle fixed vertices, like *Scotch*. This should allow us to partition much larger graph in larger part number. As another perspective, this approach can be relevant in the context of code coupling: e.g., if one code becomes more computationally intensive relatively to the other, it could be valuable to dynamically migrate some processor resources to the other code, and thus to equilibrate the whole coupled application. This work is currently conducted in the framework of Clément Vuchener PhD thesis and should be defended in september 2013.

6.3. Distributed Shared Memory approach for the steering of parallel simulations

As a different approach of EPSN, we recently propose in the thesis of J. Soumagne *an in-situ visualization approach for parallel coupling and steering of simulations through distributed shared memory files (DSM)*. Indeed, as simulation codes become more powerful and more interactive, it is desirable to monitor a simulation in-situ, performing not only visualization but also analysis of the incoming data as it is generated. Monitoring or post-processing simulation data in-situ has obvious advantage over the conventional approach of saving to – and reloading data from – the file system; the time and space it takes to write and then read the data from disk is a significant bottleneck for both the simulation and subsequent post-processing steps. Furthermore, the simulation may be stopped, modified, or potentially steered, thus conserving CPU resources.

In this thesis, we propose a loosely coupled approach that enables a simulation to transfer data to a visualization server via the use of in-memory files. We show in this study how the interface, implemented on top of a widely used hierarchical data format (HDF5), allows us to efficiently decrease the I/O bottleneck by using efficient communication and data mapping strategies. For steering, we present an interface that allows not only simple parameter changes but also complete re-meshing of grids or operations involving regeneration of field values over the entire computational domain to be carried out. This approach is generic enough so that no particular knowledge of the underlying model is required and a user can therefore plug any simulation to this framework without any re-compilation work.

A scalability study have demonstrated the performance of this solution up to 2048 cores on a Cray machine. Finally, the environment has been validated on two industrial test cases: the first one is developed by Ecole Centrale de Nantes and HydrOcean and an object placed into a wave maker is dynamically modified and steered, thereby making use of the re-meshing capabilities introduced by the framework; and the other is developed by Ecole Centrale de Lyon and ANDRITZ HYDRO, a Pelton turbine is dynamically controlled and results are analyzed in-situ.

This thesis has been defended by J. Soumagne in december 2012. This work was supported by NextMuSE project receiving funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement 225967. It has been realized in collaboration with the Swiss National Supercomputing Centre (CSCS). J.

6.4. Material physics

6.4.1. Hybrid materials

The study of hybrid materials based on a coupling between molecular dynamics (MD) and quantum mechanism (QM) simulation has been conducted in collaboration with IPREM (Pau) within the ANR CIS 2007 NOSSI (ended December 2011). These simulations are complex and costly and may involve several length scales, quantum effects, components of different kinds (mineral-organic, hydro-philic and -phobic parts). Our goal was to compute dynamical properties of hybrid materials like optical spectra. The computation of optical spectra of molecules and solids is the most consuming time in such coupling. This requires new methods designed for predicting excited states and new algorithms for implementing them. Several tracks have been investigated in the project and new results obtained as described bellow.

Optical spectra.

Some new improvements in our TD-DFT code have been introduced. Our method is based on the LCAO method for densities and excited states that computes electronic excitation spectra. We have worked in two directions:

- As the method introduces a regularization parameter to obtain regularized spectra we have used it to build better algorithms. In particular, we have developed a new hierarchical algorithm that builds a well adapted frequency distribution to better capture the biggest peaks (strongest oscillator strengths) in the spectrum. Moreover, a nonlinear fit method was added and used to compute the transitions and the oscillator strengths of the spectrum.
- In our algorithm, we used a coarse grain paradigm to parallelize the spectrum computation. This approach leads to a memory bottleneck for large systems. In that respect, we have explored a new parallel approach based on a fine grain paradigm (matrix-vector parallelization) to better exploit the manycore architecture of the emerging computers.

Finally, the code called *fast*, is released of the inria's gforge.

QM/MM algorithm. For structure studies or dynamical properties, we have coupled QM model based on pseudo-potentials (SIESTA code) with dynamic molecular (DL-POLY code). Therefore we have developed a new algorithm to avoid accounting twice for the forces and the quantum electric field in the molecular model. All algorithms involved in the coupling have been introduced both in SIESTA and in DL-POLY codes. The following new developments needed by the coupling have been introduced in the SIESTA code:

- We have implemented a fast evaluation of the molecular electrostatic field on the quantum grid.
- We have introduced a non periodic Poisson solver based on the parallel linear Hypre solver. This solver allows us to use computation domains as small as possible.
- We have implemented the ElectroStatic Potential (ESP) fit method to obtain more physical point charges than those given by SIESTA with the Mulliken method. These point charges are used by the MM codes to compute electrostatic forces.

Thanks to all our developments introduced in SIESTA a collaboration with the SIESTA research team has started. This enables us to have access to their private svn like repository. Preliminary results on a water dimer and a water box systems show good agreement with other methods developed in SIESTA and DL-POLY teams.

these results were presented in [29], [30].

6.4.2. Material failures

We have started in the context of the OPTIDIS ANR to work on dislocation simulations. The main characteristic of these simulations is that they are highly dynamical. This year, we focused on the definition of efficient cache aware data structure to manage points and segments. All the algorithms have been adapted to this structure and we have stared the development of the OPTIDIS prototype. This prototype has been parallelized with OpenMP model. More physics will be added by our partners that will give us the capability to grow our simulation and run some meaningful benchmark.

We will work in three directions. Firstly, we will investigate how to adapt our fast multipole method to compute constraints and then forces in the context of FastLA associated team. Secondly, we will improved the displacement of the segments and the way to treat collision in parallel. Finally, we will move on hybrid parallelism for our prototype.

HIPERCOM Project-Team

6. New Results

6.1. Time Slot Assignment in Wireless Sensor Networks

Participants: Pascale Minet, Ridha Soua, Erwan Livolant.

6.1.1. NP-completeness of the Time Slot Assignment problem

In data gathering applications, wireless sensor networks (WSNs) collect data from sensor nodes towards a sink in a multi-hop convergecast structure. Assigning equal channel access to each node may lead to congestion and inefficient use of the bandwidth. That is why we focus on traffic-aware solutions. More precisely, we investigate the Time Slot Assignment problem, where nodes are assigned time slots to transmit their data to the sink, while minimizing the total number of slots. We considered the generalized h-hop Time Slot Assignment problem for any positive integer h, where any two nodes that are less than or equal to h-hop away are not scheduled simultaneously. We proved its NP-completeness.

6.1.2. Multichannel Slot Assignment

The throughput requirement of data gathering applications is difficult to meet with a single wireless channel. Furthermore, the considered channel may be temporarily jammed. That is why, we focus on a multichannel time slot assignment that minimizes the data gathering cycle. We first formalize the problem as a linear program and compute the optimal time needed for a raw data convergecast in various multichannel topologies (linear, multi-line, tree). These optimal times apply to nodes equipped with one or several radio interfaces. This work generalizes the results established by Incel. We then propose our algorithm called MODESA and prove its optimality in various multichannel topologies. We evaluate its performances in terms of number of slots, maximum buffer size and number of active/sleep switches per node. Furthermore, we present variants of MODESA achieving a load balancing between the channels used.

6.1.3. Multisink Multichannel Slot Assignment

We generalize this work, taking into account the existence of several sinks. We focus on the data gathering problem with differentiated traffic, each addressed to a specific sink in multichannel WSNs. In order to find a collision-free optimized multichannel time slot assignment that minimizes the data gathering cycle, we propose a centralized traffic-aware algorithm called MUSIKA. We formulate the problem as a linear program and compute the optimal time needed for a raw data convergecast in various multichannel topologies (linear, multi-line, tree). More generally, we run simulations on various network topologies to evaluate the performance of MUSIKA in terms of cycle length, maximum buffer size and slot reuse ratio for different use cases: redundant functional processing chains, different application functionalities per sink.

6.2. Multi-Sink Wireless Sensor deployment and energy analysis

Participants: Paul Mühlethaler, Nadjib Achir.

We propose a general framework for multi-sink Wireless Sensors networks (WNSs). This framework is devoted to computing the optimal deployment of sinks for a given maximum number of hops between nodes and sinks. This framework allows an estimation of the energy consumption to be computed. We consider the energy consumed due to reporting, forwarding and overhearing. In contrast to reporting and forwarding, the energy used in overhearing is difficult to estimate because it is dependent on the packet scheduling. We determine the upper-bound and lower-bound of overhearing. We also propose another estimation which can simulate non interfering parallel transmissions which is more tractable in large networks. We note that overhearing largely predominates in energy consumption. A large part of the optimizations and computations carried out in this paper are obtained using ILP formalization.

6.3. WSN Redeployment

Participants: Pascale Minet, Saoucene Mahfoudh Ridene, Ines Khoufi.

This is a joint work with Telecom SudParis: Anis Laouiti.

6.3.1. *Centralized redeployment algorithm based on Virtual Forces*

In many applications (e.g military, environment monitoring), wireless sensors are randomly deployed in a given area. Unfortunately, this deployment is not efficient enough to ensure full area coverage and total network connectivity. Hence, all the considered area must be covered by sensors ensuring that any event is detected in the sensing range of at least one sensor. In addition, the sensor network must be connected in terms of radio communication in order to forward the detected event to the sink(s). Thus, a redeployment algorithm has to be applied in order to achieve these two goals. In this context, we have proposed redeployment algorithms based on virtual forces. First, we have designed and simulated a centralized algorithm called CVFA. This algorithm is executed by a specific node which has global information of node positions.

6.3.2. *Distributed redeployment algorithm based on Virtual Forces*

Then, we proposed DVFA, Distributed Virtual forces Algorithm. Each node in the network executes DVFA and computes its new position based on information collected from its neighbors.
Performance evaluation shows that both CVFA and DVFA give very good coverage rate (between 98% and 100%) and ensure the connectivity between sensors.

6.3.3. *Distributed redeployment algorithm based on Virtual Forces in the presence of obstacles*

Moreover, in a real environment, obstacles such as trees, walls and buildings may exist and they may impact the deployment of wireless sensors. Obstacles can prohibit the network connectivity between nodes and create some uncovered holes or some accumulation of sensors in the same region. Consequently, an efficient wireless sensors deployment algorithm is required to ensure both coverage and network connectivity in the presence of obstacles. We have focused on this problem and enhanced our Distributed Virtual Force Algorithm (DVFA) to cope with obstacles. Simulation results show that DVFA gives very good performances even in the presence of obstacles.

6.4. Mesh Network Planning: Deployment and Canal Allocation

Participant: Nadjib Achir.

This is a joint work with University Paris XIII: A. Farsi, K. Boussetta.
We deal with the Wireless LAN planning problem. We study this problem and we propose to couple its two major issues: AP placement and channel assignment to treat them jointly. Here, we propose a novel fast and scalable three-phase heuristic algorithm (TPHA). Our proposal is able to resolve the defined multiobjective problem to provide (1) the efficient number of Access Points (APs) to be deployed, while (2) ensuring the coverage of all Test Points (TPs) and (3) maximizing their nominal data rate. To achieve the first objective, we propose an heuristic called MCL-ILP combining the quick decision making based on the Markovian CLustering algorithm and the exact solution provided by the Integer Linear Programming. Hence, a TPs-based Least Interfering Channel Search algorithm (TLICS) has been proposed for channel assignment to improve the throughput at TP locations. However, the Virtual Forces-based WLAN Planning Algorithm namely VFPA considers the results delivered by the two previous algorithms as an initial solution and tries to enhance it by adjusting the APs' positions and re-assigning their operating frequencies. Computational results exhibit that our proposal is highly beneficial to designing WLANs.

6.5. Routing in MANETs using slotted Aloha. End-to-end delays

Participants: Paul Mühlethaler, Iskander Banaouas.

This is a joint work with TREC: B. Blaszczyzyn.

Planar Poisson models with the Aloha medium access scheme have already proved to be very useful in studies of mobile ad-hoc networks (MANETs). However, it seems difficult to quantitatively study the performances of end-to-end routing in these models. In order to tackle this problem, in this paper we study a *linear stationary route embedded in an independent planar field of interfering nodes*. We consider this route as an idealization of a “typical” route in a MANET obtained by some routing mechanism. Such a decoupling allows us to obtain many numerically tractable expressions for local and mean end-to-end delays and the speed of packet progression, assuming slotted Aloha MAC and the Signal-to-Interference-and-Noise Ratio (SINR) capture condition, with the usual power-law path loss model and Rayleigh fading. These expressions show how the network performance depends on the tuning of Aloha and routing parameters and on the external noise level. In particular we show a need for a well-tuned lattice structure of fixed relaying nodes, which helps to relay packets on long random routes in the presence of a non-negligible noise. We also consider a *Poisson-line MANET model*, in which *all* nodes are located on roads forming a Poisson-line process. In this case our linear route is rigorously (in the sense of Palm theory) the typical route in this Poisson-line MANET.

6.6. Cognitive networks using a darwinian approach

Participant: Paul Mühlethaler.

This is a joint work with Alcatel Bell Labs: Philippe Jacquet.
We present a new approach for cognitive radio. In the usual approach the secondary network is in charge of monitoring the channel to determine whether or not the primary network is active in the area. If it is not, the secondary network is allowed to use the spectrum. In the new access scheme we propose, the primary network encompasses the techniques which allow it to capture the bandwidth even if the secondary network is transmitting in the area. The access scheme of the primary network preempts the secondary network activity. We present an access scheme which preempts the IEEE 802.11 decentralized scheme. This protocol is a generalized Carrier Sense Multiple Access scheme using active signaling. Instead of only sensing the carrier, this algorithm also transmits bursts of signal which may be sensed by the other nodes. If so, they give up the selection process. We show that this scheme preempts the IEEE 802.11 decentralized access scheme if the bursts transmitted by the node in the primary network are made up of special sequences which alternate between bursts of signal and periods of sensing. These sequences called (d, k) sequences encompass a minimum number d and a maximum number of k successive zeros during which the node senses the channel to find other possible concurrent transmissions. In practice we use $d = 0$ and k depends on the duration of the IEEE 802.11 interframe space and the duration of a signaling burst. We compute the number of $(0, k)$ sequences with respect to the length n of the sequence. We also show that (d, k) sequences (with $2d > k$) can be used if, by mistake, during the signaling phase one burst is not detected. We evaluate the number of such sequences.

6.7. Massive mobile dense wireless networks

Participants: Aline Carneiro Viana, Ana Cristina B. Kochem Vendramin, Kanchana Thilakarathna, Eduardo Mucceli.

routing protocols, analytical models, content distribution.

6.7.1. Scientific achievements

6.7.1.1. Social Relationship Classified

Understanding human mobility is of fundamental importance when designing new communication protocols that exploit opportunistic encounters among users. In particular, human behavior is characterized by an elevated rate of regularity, but random events are always possible in the routines of individuals as hardly predictable situations that deviate from the regular pattern and are unlikely to arise repeatedly in the future. These random events veil the ordinary patterns by introducing a significant amount of noise, thus making the process of knowledge discovery in social dataset a complex task. However, the ability to accurately identify random and social events in large datasets is essential to social analysis as well as to applications that rely on a precise description of human routines, such as recommendation systems, forwarding strategies and ad-hoc message dissemination schemes focusing on coverage efficiency with a limited number of redundant messages. In such a context, we have proposed a strategy to analyze wireless network scenarios where mobile users interact in a rational manner, reflecting their interests and activity dynamics. Our strategy, named Random rElationship ClASsifier sTrategy (RECAST), allows to classify user relationships, separating random interactions from different kinds of social ties. The goal is achieved by observing how the real system differs from an equivalent one where entities decisions are completely random. We have evaluate the effectiveness of RECAST classification on datasets of real-world user contacts in diverse networking contexts. Our analysis unveils significant differences in the relationship dynamics of the datasets, proving that the evaluation of network protocols on a single dataset cannot lead to conclusions of general validity.

6.7.1.2. Social-aware Forwarding Protocol

Pervasiveness of computing devices, ubiquitous wireless communication, emergence of new applications, and cloud services are examples of current new emerging factors that emphasize the increasing need for adaptive networking solutions. The adaptation, most of the time, requires the design of more interdisciplinary approaches as those inspired by techniques coming from biology, social structures, games, and control systems. The approach we consider brings together solutions from different but complementary domains - i.e., networking, biology, and complex networks - aiming to deal with the problem of efficient data delivery in mobile and intermittently connected networks. For this, we have designed the Cultural Greedy Ant (CGrAnt) protocol to solve the problem of data delivery in mobile and intermittently connected networks referred as Delay Tolerant Networks (DTNs). CGrAnt is a hybrid Swarm Intelligence-based forwarding protocol designed to deal with the dynamic and complex environment of DTNs resulting from users mobility or varying conditions of wireless communications. CGrAnt is based on (1) Cultural Algorithms (CA) and Ant Colony Optimization (ACO) and (2) metrics which characterize opportunistic social connectivity between wireless users. CA and ACO are used to direct the network traffic, taking into account a set of social-aware metrics that may infer relevant structures in meeting regularities and mobility patterns of users. The most promising message forwarders are selected through a greedy transition rule based on local and global information captured from the DTN environment. Through simulation, we have analyzed the influence of ACO operators and CA's knowledge on CGrAnt performance. We have then compared the performance of CGrAnt with PROPHET and Epidemic protocols under varying networking parameters. Results have shown that CGrAnt achieves the highest delivery ratio and lowest byte redundancy.

6.7.1.3. Opportunistic Content Dissemination

Here, we focus on dissemination of content for delay tolerant applications/services, (i.e. content sharing, advertisement propagation, etc.) where users are geographically clustered into communities. Due to emerging security and privacy related issues, majority of users are becoming more reluctant to interact with strangers and are only willing to share information/content with the users who are previously identified as friends. In this environment, opportunistic communication will not be effective due to the lack of known friends within the communication range. Thus, we have proposed a novel architecture that addresses the issues of lack of trust, timeliness of delivery, loss of user control, and privacy-aware distributed mobile social networking by combining the advantages of distributed decentralized storage and opportunistic communications. We have formally defined a content replication problem in mobile social networks and show that it is computationally hard to solve optimally. Then, we have proposed a community based greedy heuristic algorithm with novel dynamic centrality metrics to replicate content in well-selected users, to maximize the content dissemination with limited number of replication. Using both real world and synthetic traces, we have shown that content replication can attain a large coverage gain and reduce the content delivery latency.

6.7.1.4. Data Offloading-aware Hotspot Deployment

With the steady growth of sales of smart-phones, the demand for services that generate mobile data traffic has grown tremendously. The growing use of traffic data generated from mobile devices overloads the network infrastructure, which is not always prepared to receive such demand. To tackle this problem, we are studying the mobile behavior and resource consumptions of people on a metropolitan area in a major city and turn it into a set of well located WiFi hotspots. For this, we have proposed a data offloading-aware hotspot deployment. It is methodologically divided as (i) creation of a time dependent weighted graph to represent people's mobility, traffic and its relation with places/locations able to receive a hotspot, (ii) measurement of location's importance and selection of the best-ranked ones. Better positioned hotspots are likely to provide better coverage, and therefore, be able to offload more data.

6.7.2. Collaborations

- Professors Anelise Munaretto and Myriam Regattieri Delgado from Federal Technological University of Parana (UTFPR), Brazil,
- Professors Aruna Seneviratne and Henrik Petander from NICTA and School of EE&T, UNSW, Sydney, Australia,
- Pedro O.S. Vaz de Melo and Antonio A. F. Loureiro, Federal University of Minas Gerais, Brazil,
- Marco Fiore and Frederic Le Mouel from INSA Lyon, France,
- Katia Jaffrès-Runser, University of Toulouse, IRIT/ENSEEIHT, France.

6.8. New services and protocols

Participants: Aline Carneiro Viana, Guilherme Maia.

6.8.1. Scientific achievements

6.8.1.1. Network Discovery

Network discovery is a fundamental task in different scenarios of IEEE 802.15.4-based wireless personal area networks. Scenario examples are body sensor networks requiring health- and wellness-related patient monitoring or situations requiring opportunistic message propagation. Therefore, we have investigated optimized discovery of IEEE 802.15.4 static and mobile networks operating in multiple frequency bands and with different beacon intervals. We designed a linear programming model that allows finding two optimized strategies, named OPT and SWOPT, to deal with the asynchronous and multi-channel discovery problem. We have also proposed a simplified discovery solution, named SUBOPT, featuring a low-complexity algorithm requiring less memory usage. A cross validation between analytical, simulation, and experimental evaluation methods was performed. Our performance studies confirmed improvements achieved by our solutions in terms of first, average, and last discovery time as well as discovery ratio, when compared to IEEE 802.15.4 standard approach and the SWEEP approach known from the literature.

6.8.1.2. Distributed Data Storage

The deployment of large-scale Wireless Sensor Network (WSN) applications (e.g., environment sensing and military surveillance), which operate unattended for long periods of time and generate a considerable amount of data, poses several challenges. One of them is *how to retrieve the sensed data*. To tackle this issue, we have designed ProFlex, a distributed data storage protocol for large-scale heterogeneous wireless sensor networks (HWSNs) with mobile sinks. ProFlex guarantees robustness in data collection by intelligently managing data replication among selected storage nodes in the network. Contrarily to related protocols in the literature, ProFlex considers the resource constraints of sensor nodes and constructs multiple data replication structures, which are managed by more powerful nodes. Additionally, ProFlex takes advantage of the higher communication range of such powerful nodes and uses the long-range links to improve data distribution by storage nodes. When compared with related protocols, we have shown through simulation that Proflex has an acceptable performance under message loss scenarios, decreases the overhead of transmitted messages, and decreases the occurrence of the energy hole problem. Moreover, we have proposed an improvement that allows the protocol to leverage the inherent data correlation and redundancy of wireless sensor networks in order to decrease even further the protocol's overhead without affecting the quality of the data distribution by storage nodes.

6.8.2. Collaborations

- PhD Niels Karowski, Technische Universitat Berlin, Germany,
- Professor Adam Wolisz, Technische Universitat Berlin, Germany,
- Antonio A. F. Loureiro, Federal University of Minas Gerais, Brazil,

INDES Project-Team

6. New Results

6.1. Security

Participants: Ilaria Castellani, Zhengqin Luo, Tamara Rezk [correspondant], José Santos, Manuel Serrano.

6.1.1. Session types with security

We have pursued our work on integrating security constraints within session types, in collaboration with our colleagues from Torino University. This resulted in the journal paper [8]. This article extends a previous conference paper with full proofs, additional examples and further results. In particular, [8] presents new properties of information-flow security, which is stronger and more compositional (*i.e.*, more robust with respect to parallel composition of processes) than that originally proposed, while being still ensured by the same session type system.

All the work on session types was partially funded by the ANR-08- EMER-010 grant PARTOUT. It is expected to continue within the starting COST Action BETTY.

6.1.2. Mashic Compiler: Mashup Sandboxing Based on Inter-frame Communication

Mashups are a prevailing kind of web applications integrating external gadget APIs often written in the Javascript programming language. Writing secure mashups is a challenging task due to the heterogeneity of existing gadget APIs, the privileges granted to gadgets during mashup executions, and Javascript's highly dynamic environment.

We propose a new compiler, called Mashic, for the automatic generation of secure Javascript-based mashups from existing mashup code. The Mashic compiler can effortlessly be applied to existing mashups based on a wide-range of gadget APIs. It offers security and correctness guarantees. Security is achieved by using the Same Origin Policy. Correctness is ensured in the presence of benign gadgets, that satisfy confidentiality and integrity constraints with regard to the integrator code. The compiler has been successfully applied to real world mashups based on Google maps, Bing maps, YouTube, and Zwibbler APIs.

This work appeared in CSF'12 [14]. See also software section.

6.1.3. A Certified Lightweight Non-Interference Java Bytecode Verifier

We propose a type system to verify the non-interference property in the Java Virtual Machine. We verify the system in the Coq theorem prover. This work will appear in the journal of Mathematical Structures in Computer Science [6].

6.2. Models, semantics, and languages

Participants: Pejman Attar, Gérard Berry, Gérard Boudol, Frédéric Boussinot, Ilaria Castellani, Johan Grande, Cyprien Nicolas, Tamara Rezk, Manuel Serrano [correspondant].

6.2.1. Memory Models

As regards the theory of multithreading, we have extended our operational approach to capture more relaxed memory models than simple write buffering. A step was made in this direction by formalizing the notion of a speculative computation, but this was not fully satisfactory as an operational approach to the theory of memory models: indeed, in the speculative framework one has to reject a posteriori some sequences of executions as invalid. In [13] we have defined a truly operational semantics, by means of an abstract machine, for extremely relaxed memory models like the one of PowerPC. In our new framework, the relaxed abstract machine features a "temporary store" where the memory operations issued by the threads are recorded, in program order. A memory model then specifies the conditions under which a pending operation from this sequence is allowed to be globally performed, possibly out of order. The memory model also involves a "write grain," accounting for architectures where a thread may read a write that is not yet globally visible. Our model is also flexible enough to account for a form of speculation used in PowerPC machines, namely branch prediction. To experiment with our framework, we found it useful to design and implement a simulator that allows us to exhaustively explore all the possible relaxed behaviors of (simple) programs. The main problem was to tame the combinatory explosion due to the massive non-deterministic interleaving of the relaxed semantics. Introducing several optimizations described in [13], we were able to run a large number of litmus tests successfully.

6.2.2. Dynamic Synchronous Language with Memory

We have investigated the language DSLM (Dynamic Synchronous Language with Memory), based on the synchronous reactive model. In DSLM, systems are composed of several sites, each of which runs a number of agents. An agent consists of a memory and a script. This script is made of several parallel components which share the agent's memory. A simple form of migration is provided: agents can migrate from one site to another. Since sites have different clocks, a migrating agent resumes execution at the start of the next instant in the destination site. Communication between a migrating agent and the agents of the destination site occurs via (dynamically bound) events. The language uses three kinds of parallelism: 1) synchronous, cooperative and deterministic parallelism among scripts within an agent, 2) synchronous, nondeterministic and confluent parallelism among agents within a site, and 3) asynchronous and nondeterministic parallelism among sites. Communication occurs via both shared memory and events in the first case, and exclusively via events in the other two cases. Scripts may call functions or modules which are handled in a host language. Two properties are assured by DSLM: reactivity of each agent and absence of data-races between agents. Moreover, the language offers a way to benefit from multi-core and multi-processor architectures, by means of the notion of synchronized scheduler which abstractly models a computing resource. Each site may be expanded and contracted dynamically by varying its number of synchronized schedulers. In this way one can model the load-balancing of agents over a site.

A secure extension of the language DSLM, called DSSLM (Dynamic Secure Synchronous Language with Memory), is currently under investigation. This language uses the same deterministic parallel operator for scripts as DSLM. It adds to DSLM a let operator that assigns a security level to the defined variable. Security levels are also assigned to events and sites, to allow information flow control during interactions and migrations. The study of different security properties (both sensitive and insensitive to the passage of the instants) and of type systems ensuring these properties is currently under way.

6.2.3. jthread

The jthread library (working name) is a Bigloo library featuring threads and mutexes and most notably a deadlock-free locking primitive. The jthread library appears as an alternative to Bigloo's pthread (POSIX threads) library and relies on it for its implementation.

The locking primitive is the following: `(synchronize* ml [:prelock mlp] expr1 expr2 ...)` where ml and mlp are lists of mutexes.

This primitive evaluates the expressions that constitute its body after having locked the mutexes in ml and before unlocking them back. The meaning of the *prelock* argument is to be explained below.

The absence of deadlocks is guaranteed by two complementary mechanisms:

- Each mutex belongs to a *region* defined by the programmer. Regions form a lattice which is inferred at runtime. A thread owning a mutex belonging to region R0 can only lock a mutex belonging to region R1 if R1 is lower than R0 in the lattice. This rule is enforced at runtime and guarantees the absence of deadlocks involving mutexes belonging to different regions.
- Under the previous condition, a thread owning a mutex M1 can lock a mutex M2 belonging to the same region only provided that M2 appeared in the *prelock* list of the `synchronize*` that locked M1. This rule is enforced at runtime and allows a *deadlock-avoiding* scheduling of threads based on previous work by Gérard Boudol and on Lamport's Bakery algorithm.

The library has been implemented. It is currently being integrated to Bigloo and benchmarked. It has not been released yet.

6.3. Web programming

Participants: Zhengqin Luo, Cyprien Nicolas, Tamara Rezk, Bernard Serpette, Manuel Serrano [correspondant].

6.3.1. Reasoning about Web Applications: An Operational Semantics for HOP

We propose a small-step operational semantics to support reasoning about web applications written in the multi-tier language HOP. The semantics covers both server side and client side computations, as well as their interactions, and includes creation of web services, distributed client-server communications, concurrent evaluation of service requests at server side, elaboration of HTML documents, DOM operations, evaluation of script nodes in HTML documents and actions from HTML pages at client side. We also model the browser same origin policy (SOP) in the semantics. We propose a safety property by which programs do not get stuck due to a violation of the SOP and a type system to enforce it. This work appeared in TOPLAS [7].

6.3.1.1. Hiphop

We pursued the development of the Hiphop orchestration language. The first version was written as a DSL with very few connection to Hop. During this year, we changed Hiphop syntax to blend it better with Hop. All Hiphop objects are now Hop values, and thus Hiphop programs can benefit from all Hop featues. The Hiphop development has enabled us to improve Hop stability and quality in client code generation.

We have found a new use-case for Hiphop: Robotics. We are currently working with the Inria Coprin team to pilot their robot using Hop and Hiphop. We have already used Hop to program with low-level motors API (using the Phidget libraries). Hop enabled us to distribute the robot control application over HTTP, in order to control the robot from a smartphone or tablet.

6.3.2. A CPS definition of HipHop

Since the Esterel model is used very dynamically in the HipHop framework, we have begun studying new frameworks of computations. We designed a definition of a HipHop-core, which is similar to Esterel-core, based on continuations. This approach allows a specification close to the implementation. The main problem was to define a predicate assuring the *absence* of a specific signal in the current instant. For this, we have designed a static analysis that predicts, for each program point and for each signal, the number of emissions remaining to be done until the end of the instant. The prediction may be over-estimated but when a null value is reached the corresponding signal can be considered as absent for the analyzed instant.

Contrary to existing analyses, this prediction can be done at compile time. Nevertheless some extra computations must be inserted in the evaluator to adjust a runtime prediction. For example. this is done when one branch of a conditional is dynamically taken, but adjusting the prediction only involves subtractions on global counters.

The continuation based definition doesn't prevent a space efficiency implementation. Esterel is known to be compiled to hardware and thus able to run a program in a fixed space of silicon; in the same manner we have implemented an evaluator than doesn't allocate extra memories while running a program: all the continuations can be allocated at compile time.

We have also extended the language to reach the HipHop definition. Some dynamic extensions (`mappar`) may dynamically allocate some resources but we were able to tune the static analysis to insure both confluence and constructive absence detection.

KERDATA Project-Team

6. New Results

6.1. Optimizing MapReduce processing

6.1.1. Hybrid infrastructures

Participants: Alexandru Costan, Bharath Vissapragada, Gabriel Antoniu.

As Map-Reduce emerges as a leading programming paradigm for data-intensive computing, today's frameworks which support it still have substantial shortcomings that limit its potential scalability. At the core of Map-Reduce frameworks stays a key component with a huge impact on their performance: the storage layer. To enable scalable parallel data processing, this layer must meet a series of specific requirements. An important challenge regards the target execution infrastructures. While the Map-Reduce programming model has become very visible in the cloud computing area, it is also subject to active research efforts on other kinds of large-scale infrastructures, such as desktop grids. We claim that it is worth investigating how such efforts (currently done in parallel) could converge, in a context where large-scale distributed platforms become more and more connected together.

In 2012 we investigated several directions where there is room for such progress: they concern storage efficiency under massive data access concurrency, scheduling, volatility and fault-tolerance. We placed our discussion in the perspective of the current evolution towards an increasing integration of large-scale distributed platforms (clouds, cloud federations, enterprise desktop grids, etc.) ([16]). We proposed an approach which aims to overcome the current limitations of existing Map-Reduce frameworks, in order to achieve scalable, concurrency-optimized, fault-tolerant Map-Reduce data processing on hybrid infrastructures. We are designing and implementing our approach through an original architecture for scalable data processing: it combines two approaches, BlobSeer and BitDew, which have shown their benefits separately (on clouds and desktop grids respectively) into a unified system. The global goal is to improve the behavior of Map-Reduce-based applications on the target large-scale infrastructures. The internship of Bharath Vissapragada was dedicated to this topic.

This approach will be evaluated with real-life bio-informatics applications on existing Nimbus-powered cloud testbeds interconnected with desktop grids.

6.1.2. Scheduling: Maestro

Participants: Shadi Ibrahim, Gabriel Antoniu.

As data-intensive applications became popular in the cloud, data-intensive cloud systems call for empirical evaluations and technical innovations. We have investigated some performance limits in current MapReduce frameworks (Hadoop in particular). Our studies reveal that the current Hadoop's scheduler for map tasks is inadequate, as it disregards replicas distributions. It causes performance degradation due to a high number of non-local map tasks, which in turn causes too many needless speculative map tasks and leads to imbalanced execution of map tasks among data nodes. We addressed these problems by developing a new map task scheduler called Maestro.

In [19], we developed a scheduling algorithm (Maestro) to alleviate the nonlocal map tasks executions problem of MapReduce. Maestro is conducive to improving the locality of map tasks executions efficiency by virtue of the finer-grained replica aware execution of map tasks, thereby having one additional factor for the chunk hosting status: the expected number of map tasks executions to be launched. Maestro keeps track of the chunks' locations along with their replicas' locations and the number of other chunks hosted by each node. In doing so, Maestro can efficiently schedule the map task to the node with minimal impacts on other nodes' local map tasks executions. Maestro schedules the map tasks in two waves: first, it fills the empty slots of each data node based on the number of hosted map tasks and on the replication scheme for their input data; second, runtime scheduling takes into account the probability of scheduling a map task on a given machine depending on the replicas of the task's input data. These two waves lead to a higher locality in the execution of map tasks and to a more balanced intermediate data distribution for the shuffling phase.

We evaluated Maestro through a set of experiments on the Grid'5000 [35] testbed. Preliminary results [19] show the efficiency and scalability of our proposals, as well as additional benefits brought forward by our approach.

6.1.3. Fault tolerance

Participants: Bunjamin Memishi, Shadi Ibrahim, Gabriel Antoniu.

The simple philosophy of MapReduce has made huge community interest for its exploration, especially in environments where data-intensive applications are primary concern. Fault tolerance is one of the key features of the MapReduce system. MapReduce tasks are re-executed in case of failure, and a potential failure of a single master causes an additional bottleneck. It is observed that the detection of the failed worker tasks in Hadoop have a certain delay, yet not solved. Willing to improve the applications performance and optimal resource utilization, both of this concerns were more than a motivation so that we show in [36] that a little attention has been devoted to the failure detection in Hadoop's MapReduce which currently uses a timeout based mechanism for detecting failed tasks.

We have performed an in-depth analysis of MapReduce's failure detection, and these preliminary studies have revealed that the current static timeout value (600 seconds) is not adequate and demonstrate significant variations in the application's response time with different timeout value. Moreover, in the presence of single machine failure, the applications latencies vary not only in accordance to the occupancy time of the failure, similar to [33], but also vary with the job length (short or long).

Based on our aforementioned micro-analysis of failure detection in MapReduce, we are currently investigating an adaptive failure detection mechanism for Hadoop, which basically addresses the timeout adjustment in real-time for different jobs and applications, so that finally to adjust this model into a Shared Hadoop Cluster. Another work should discuss in details different failures types in MapReduce system and survey the different mechanisms used in MapReduce for detecting, handling and recovering from these failures and their inherited pros and cons; additionally, to a particular interest will be the analyzing of different execution environments including Cluster, Cloud and Desktop Grid on the efficiency of fault-tolerance in MapReduce. This work will soon be published.

6.2. A-Brain and TomusBlobs

6.2.1. TomusBlobs

Participants: Radu Tudoran, Alexandru Costan, Gabriel Antoniu.

Enabling high-throughput massive data processing on cloud data becomes a critical issue, as it impacts the overall application performance. In the framework of the MSR-Inria A-Brain co-led by Gabriel Antoniu (KerData) and Bertrand Thirion (PARIETAL), the TomusBlobs[22] system was designed and implemented by KerData to address such challenges at the level of the cloud storage. The system we introduce is a concurrency-optimized data storage system which federates the virtual disks associated to VMs. As TomusBlobs does not require modifications to the cloud middleware, it can serve as a high-throughput globally-shared data storage for the cloud applications that require data passing among computation nodes.

We leveraged the performance of this solution to enable efficient data-intensive processing on commercial clouds by building an optimized prototype MapReduce framework for Azure. The system, deployed on 350 cores in Azure, was used to execute a real-life application, A-Brain with the goal of searching for significant associations between brain locations and genes.

The achieved throughput increased with an order of 2 for reading, respectively 3 for writing compared to the remote storage. With our approach for MapReduce data processing, the computation time is reduced to 50 % compared to the existing solutions, while the cost is reduced up to 30 %.

6.2.2. Iterative MapReduce

Participants: Radu Tudoran, Alexandru Costan, Gabriel Antoniu, Louis-Claude Canon.

While MapReduce has arisen as a major programming model for data analysis on clouds, there are many scientific applications that require processing patterns different from this paradigm. As such, reduce-intensive algorithms are becoming increasingly useful in applications such as data clustering, classification and mining. These algorithms have a common pattern: data are processed iteratively and aggregated into a single final result. While in the initial MapReduce proposal the reduce phase was a simple aggregation function, recently an increasing number of applications relying on MapReduce exhibit a reduce-intensive pattern, that is, an important part of the computations are done during the reduce phase. However, platforms like MapReduce or Dryad lack built-in support for reduce-intensive workloads.

To overcome these issues, we introduced MapIterativeReduce [23], a framework which: 1) extends the MapReduce programming model to better support reduce-intensive applications by exploiting the inherent parallelism of the reduce tasks which have an associative and/or commutative operation; and 2) substantially improves their efficiency by eliminating the implicit barrier between the Map and the Reduce phase. We showed how to leverage this architecture for scientific applications by enhancing the fault tolerance support in Azure and TomusBlobs, the underlying storage system, with a light checkpointing scheme and without any centralized control.

We evaluated MapIterativeReduce on the Microsoft Azure cloud with synthetic benchmarks and with a real-life application. Compared to state-of-art solutions, our approach enables faster data processing, by reducing the execution times by up to 75 %.

6.2.3. Adaptive file management for clouds

Participants: Radu Tudoran, Alexandru Costan, Gabriel Antoniu.

Recently, there is an increasing interest to execute general data processing schemas in clouds, as it would allow many scientific applications to migrate to this computing infrastructures. The natural way to do this is to designe and adopt Workflow Processing engines built for clouds. Such workflow processing in clouds would involve data propagation on the computation nodes based on well defined data access patterns. Having an efficient file management backend for a workflow engines is thus essential as we move to the world of BigData.

We proposed a new approach for a transfer-optimized file management in clouds On the one hand, our solution manages files within the deployment leveraging data locality. On the other hand, we envision an adaptive system that adopts the transfer method most suited based on the data transfer context.

The performance evaluation showed significant gains in terms of transfer throughput and computation time. File transfer times are reduced up to a factor of 5 with respect to the remote storage, while the timespan of running applications is reduced by more than 25% compared with other frameworks like Hadoop on Azure. This work was done in the context of a 3-month internship of Radu Tudoran hosted by the Advance Technology Lab from Microsoft Europe, Germany, Aachen.

6.3. Autonomic Cloud data storage management

Participants: Gabriel Antoniu, Alexandru Costan.

Providing the users with the possibility to store and process data on externalized, virtual resources from the cloud requires simultaneously investigating important aspects related to security, efficiency and quality of service. To this purpose, it clearly becomes necessary to create mechanisms able to provide feedback about the state of the storage system along with the underlying physical infrastructure. This information thus monitored, can further be fed back into the storage system and used by self-managing engines, in order to enable an autonomic behavior, possibly with several goals such as self-configuration, self-optimization, or self-healing. Within the DataCloud@work Associate Team in partnership with Politehnica University of Bucharest, our goal was to bring substantial contributions in this direction by leveraging previous efforts materialized through the BlobSeer data-sharing platform and several large-scale applications.

6.3.1. Evaluating BlobSeer for sharing application data on IaaS cloud infrastructures

. We showed how several types of large scale applications (e.g. scientific data aggregation, context-aware data management, video and image processing) rely on BlobSeer's support for high concurrency and increased data access throughput in order to achieve their goals. Several building blocks were implemented to address all the applications' requirements (new meta-data management, extended clients). An illustrative class of applications is represented by the context-aware ones. Our goal was to provide a cloud-based storage layer for sensitive context data, collected from a vast amount of sources: from smartphones to sensors located in the environment. We developed a layer on top of BlobSeer to allow two major things: efficient access to data based on meta-information (a catalogue of context data), and the support from mobility in the form of distributed caches able to support the movement of people and give support for fast access to real-time event of interest (dissemination of events of interest). The system as a whole was evaluated in extensive experiments, involving thousands of simulated clients, and the results proved its valuable contribution to advance the current state-of-the-art in the area of interested (middlewares to support context-aware apps).

6.3.2. Fault-tolerant VM management in Clouds, using BlobSeer

. We were also concerned about the fault tolerance support for the aforementioned applications on the cloud. A first step towards this goal consisted in exploring ways to deploy, boot and terminate VMs very quickly, enabling cloud users to exploit elasticity to find the optimal trade-off between the computational needs (number of resources, usage time) and budget constraints. We built a VM management system based on the FUSE interface leveraging the high throughput under increased concurrency of BlobSeer. We integrated it within the Nimbus cloud to allow fast VM deployment / snapshotting/ live migration. An adaptive prefetching mechanism is used to reduce the time required to simultaneously boot a large number of VM instances on clouds from the same initial VM image (multi-deployment). This proposal does not require any foreknowledge of the exact access pattern. It dynamically adapts to it at run time, enabling the slower instances to learn from the experience of the faster ones. Since all booting instances typically access only a small part of the virtual image along almost the same pattern, the required data can be pre-fetched in the background. In parallel, we investigated ways to ensure the anonimity of the data management layer, a requirement for HPC applications deployed into the clouds.

6.4. Advanced techniques for scalable cloud storage

6.4.1. Adaptive consistency

Participants: Houssem-Eddine Chihoub, Shadi Ibrahim, Gabriel Antoniu.

In just a few years cloud computing has become a very popular paradigm and a business success story, with storage being one of the key features. To achieve high data availability, cloud storage services rely on replication. In this context, one major challenge is data consistency. In contrast to traditional approaches that are mostly based on strong consistency, many cloud storage services opt for weaker consistency models in order to achieve better availability and performance. This comes at the cost of a high probability of stale data being read, as the replicas involved in the reads may not always have the most recent write. In [17], we propose a novel approach, named Harmony, which adaptively tunes the consistency level at run-time according to the application requirements. The key idea behind Harmony is an intelligent estimation model of stale reads, allowing to elastically scale up or down the number of replicas involved in read operations to maintain a low (possibly zero) tolerable fraction of stale reads. As a result, Harmony can meet the desired consistency of the applications while achieving good performance. We have implemented Harmony and performed extensive evaluations with the Cassandra cloud storage on Grid'5000 testbed and on Amazon EC2. The results show that Harmony can achieve good performance without exceeding the tolerated number of stale reads. For instance, in contrast to the static eventual consistency used in Cassandra, Harmony reduces the stale data being read by almost 80%. Meanwhile, it improves the throughput of the system by 45% while maintaining the desired consistency requirements of the applications when compared to the strong consistency model in Cassandra.

While most optimizations efforts for consistency management in the cloud focus on how to provide adequate trade-offs between consistency guarantees and performance, a little work has been investigating the impact of consistency on monetary cost. However, and since strict strong consistency is not always required for large class of applications, in [25] we argue that monetary cost should be taken into consideration when evaluating or selecting a consistency level in the cloud. Accordingly, we define a new metric called consistency-cost efficiency. Based on this metric, we present a simple, yet efficient economical consistency model, called Bismar, that adaptively tunes the consistency level at run-time in order to reduce the monetary cost while simultaneously maintaining a low fraction of stale reads. Experimental evaluations with the Cassandra cloud storage on a Grid'5000 testbed show the validity of the metric and demonstrate the effectiveness of the proposed consistency model allowing up to 31 % of money saving while tolerating a very small fraction of stale reads.

6.4.2. In-memory data management

Participants: Viet-Trung Tran, Gabriel Antoniu, Luc Bougé.

As a result of continuous innovation in hardware technology, computers are made more and more powerful than their prior models. Modern servers nowadays can possess large main memory capability that can size up to 1 Terabytes (TB) and more. As memory accesses are at least 100 times faster than disk, keeping data in main memory becomes an interesting design principle to increase the performance of data management systems. We design DStore [27], a document-oriented store residing in main memory to fully exploit high-speed memory accesses for high performance. DStore is able to scale up by increasing memory capability and the number of CPU-cores rather than scaling horizontally as in distributed data-management systems. This design decision favors DStore in supporting fast and atomic complex transactions, while maintaining high throughput for analytical processing (read-only accesses). This goal is (to our best knowledge) not easy to achieve with high performance in distributed environments.

To achieve its goals, DStore is built with several design principles. DStore follows a single threaded execution model to execute update transactions sequentially by one *master thread* while relying on a versioning concurrency control to enable multiple *reader threads* running simultaneously. DStore builds indexes for fast document lookups. Those indexes are built using the *delta-indexing* and *bulk updating* mechanisms for faster indexes maintenance and for atomicity guarantees of complex queries. Moreover, DStore is designed to favor stale reads that only need to access isolated snapshots of the indexes. Thus, it can eliminate interference between transactional processing and analytical processing.

We conducted multiple synthetic benchmarks on the Grid'5000 to evaluate the DStore prototype. Our preliminary results demonstrated that DStore achieved high performance even in scenarios where *Read*, *Insert* and *Delete* queries were performed simultaneously. In fact, the processing rate measured was about 600,000 operations per second for each concurrent process.

6.4.3. Scalable geographically distributed storage systems

Participants: Viet-Trung Tran, Gabriel Antoniu, Luc Bougé.

To build a globally scalable distributed file system that spreads over a wide area network (WAN), we propose an integrated architecture for a storage system relying on a distributed metadata-management system and BlobSeer, a large-scale data-management service. Since BlobSeer was initially designed to run on cluster environments, it is necessary to extend BlobSeer in order to take into account the latency hierarchy on geographically distributed environments.

We proposed BlobSeer-WAN, an extension of BlobSeer optimized for geographically distributed environments. First, in order to keep metadata I/O local to each site as much as possible, we proposed an asynchronous metadata replication scheme at the level of metadata providers. As metadata replication is asynchronous, we guarantee a minimal impact on the writing clients that generate metadata. Second, we introduced a distributed version management in BlobSeer-WAN by leveraging an implementation of multiple version managers and using vector clocks for detection and resolution of collision. This extension to BlobSeer keeps BLOBs consistent while they are globally shared among distributed sites under high concurrency.

Several experiments were performed on the Grid'5000 testbed demonstrated that BlobSeer-WAN can offer scalable aggregated throughput when concurrent clients append to one BLOB. The aggregated throughput reached to 1400 MB/s for 20 concurrent clients. We also compared BlobSeer-WAN and the original BlobSeer in local site accesses. The experiments shown that the overhead of the multiple version managers implementation and the metadata replication scheme in BlobSeer-WAN is minimal, thanks to our asynchronous replication scheme.

6.5. Scalable I/O for HPC

6.5.1. Damaris and HPC visualization

Participants: Matthieu Dorier, Gabriel Antoniu.

In the context of the Joint Inria/UIUC/ANL Laboratory for Petascale computing (JLCP), have proposed the Damaris approach to enable efficient I/O, data analysis and visualization at ver large scale from SMP machines. The I/O bottlenecks already present on current petascale systems as well as the amount of data written by HPC applications force to consider new approaches to get insights from running simulations. Trying to bypass the storage or drastically reducing the amount of data generated will be of outmost importance for exascale. In-situ visualization has therefor been proposed to run analysis and visualization tasks closer to the simulation, as it runs.

The first results obtained with Damaris in achieving scalable, jitter-free I/O, were published this year [18]. In order to achieve efficient in-situ visualization at extreme scale, we investigated the limitations of existing in-situ visualization software and proposed to fill the gaps of these software by providing in-situ visualization support to Damaris. The use of Damaris on top of existing visualization packages allows us to:

- Reduce code instrumentation to a minimum in existing simulations,
- Gather the capabilities of several visualization tools to offer adaptability under a unified data management interface,
- Use dedicated cores to hide the run time impact of in-situ visualization and
- Efficiently use memory through a shared-memory-based communication model.

Experiments are now being conducted on BlueWaters (Cray XK6 at NCSA), Intrepid (BlueGene/P at ANL) and Grid5000 with representative visualization scenarios for the CM1 [31] atmospheric simulation and the Nek5000 [34] CFD solver.

Results will be submitted to a conference in early 2013. We plan to further investigate the role that Damaris can take in performing efficient and self-adaptive data analysis in HPC simulations.

6.5.2. Advanced I/O and Storage

Participants: Matthieu Dorier, Alexandru Costan, Gabriel Antoniu.

The recent extension of the JLPC to Argonne National Lab (ANL) has opened new research directions in the field of advanced I/O and storage for HPC, in collaboration with Robert Ross's team at ANL's Mathematics and Computer Science Division (MCS). A founding from the FACCTS program (France And Chicago CollaboraTing in Science) allowed multiple visits (see Section 8.4) of students and researchers from both sides to initiate this new collaboration and explore potential research directions.

One outcome of these visits has been the adaptation of Damaris to work on BlueGene/P and BlueGene/Q machines installed at ANL. Several exchanges led to the design of new I/O scheduling algorithms leveraging Damaris for efficient asynchronous I/O and storage. These algorithms are currently being evaluated, and expected to be published in early 2013.

During these exchanges we also investigated new storage architectures for Exascale systems leveraging BLOB-based large-scale storage able to cope with complex data models. We will explore how we can combine the benefits of the approaches to Big Data storage currently developed by the partners: the BlobSeer approach (KerData), which provides support for multi- versioning and efficient fine-grain access to huge data under heavy concurrency and the Triton approach (ANL), which introduces new object storage semantics. The final goal of the resulting architecture will be to propose efficient solutions to data-related bottlenecks in Exascale HPC systems.

LOGNET Team

5. New Results

5.1. A Backward-Compatible Protocol for Inter-routing over Heterogeneous Overlay Networks

Participants: Giang Ngo Hoang, Luigi Liquori, Vincenzo Ciancaglini, Petar Maksimovic, Hung Nguyen Chan [HUST, Vietnam].

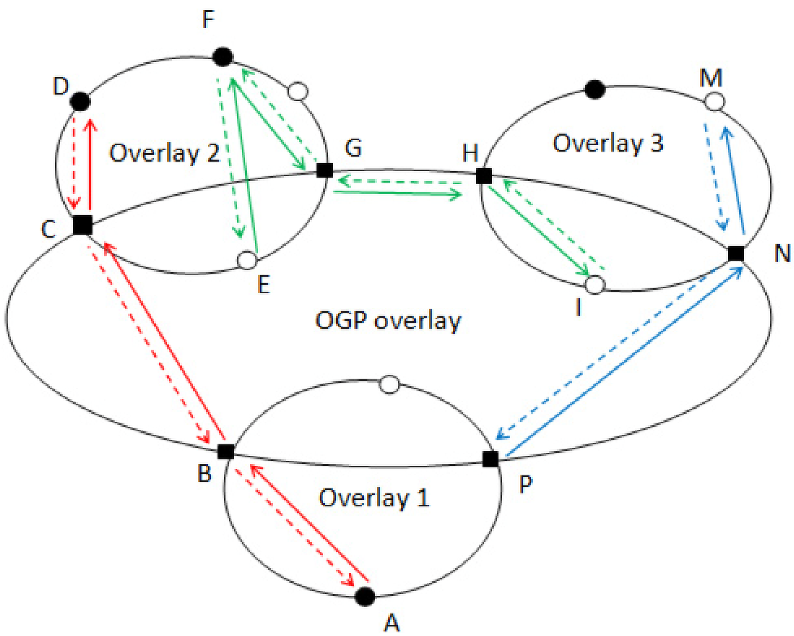


Figure 11. An Overlay Gateway Protocol Topology

Overlay networks are logical networks running on the highest level of the OSI stack: they are applicative networks used by millions of users everyday. In many scenarios, it would be desirable for peers belonging to overlays running different protocols to communicate with each other and exchange certain information. However, due to differences in their respective protocols, this communication is often difficult or even impossible to be achieved efficiently, even if the overlays are sharing common objectives and functionalities. In this paper, we address this problem by presenting a new overlay protocol, called OGP (Overlay Gateway Protocol), allowing different existing networks to route messages between each other in a backward-compatible fashion, by making use of specialized peers joined together into a super-overlay. Experimental results on a large scale Grid5000 infrastructure show that having only a small number of nodes running the OGP protocol is sufficient for achieving efficient routing between heterogeneous overlay networks.

The three scenarios in Figure 11 are shown to illustrate the routing of three lookup queries, in which full OGP peers, lightweight OGP peers and blind peers interact in order to reach across overlays represent requests, while dashed lines represent responses. using the OGP super-overlay. The three smaller ovals represent standard overlays, while the largest oval represents the OGP super-overlay, forwarding messages back and forth between standard overlays. The black squares B; C; G; N and P represent full OGP peers, the black circles A; D and F represent lightweight OGP peers, while the white circles E; H, and M represent blind peers. Solid lines requests, while dashed lines represent responses. The paper is the continuation of the work of HotPost 2011 [7] and it has been accepted to ACM SAC 2013 [33] and a long version will be submitted in a high level conference [34].

5.2. Interconnection of large scale unstructured P2P networks: modeling and analysis

Participants: Rossano Gaeta [Univ. Turin], Riccardo Loti, Luigi Liquori, Vincenzo Ciancaglini [contact].

Interconnection of multiple P2P networks has recently emerged as a viable solution to increase system reliability and fault-tolerance as well as to increase resource availability. In this paper we consider interconnection of large scale unstructured P2P networks by means of special nodes (called *Synapses*) that are co-located in more than one overlay. Synapses act as *trait d'union* by sending/forwarding a query to all the P2P networks they belong to. Modeling and analysis of the resulting interconnected system is crucial to design efficient and effective search algorithms and to control the cost of interconnection. To this end, we develop a generalized random graph based model that is validated against simulations and it is used to investigate the performance of search algorithms for different interconnection costs and to provide some insight in the characteristics of the interconnection of a large number of P2P networks. To overcome this strong limitation, we develop a generalized random graph based model to represent the topology of one unstructured P2P network, the partition of nodes into Synapses, the probabilistic flooding based search algorithms, and the resource popularity. We validate our model against simulations and prove that its predictions are reliable and accurate. We use the model to investigate the performance and the cost of different search strategies in terms of the probability of successfully locating at least one copy of the resource and the number of queries as well as the interconnection cost. We also gain interesting insights on the dependency between interconnection cost and statistical properties of the distribution of Synapses. Finally, we show that thanks to our model we can analyze the performance of a system composed of a large number of P2P networks.

To the best of our knowledge, this is the first paper on model-based analysis of interconnection of large scale unstructured P2P networks [27], [28]

5.3. SIEVE: a distributed, accurate, and robust technique to identify malicious nodes in data dissemination on MANET

Participants: Rossano Gaeta [Univ. Turin], Riccardo Loti [contact], Marco Grangetto [Univ Turin].

We consider the following problem: nodes in a MANET must disseminate data chunks using rateless codes but some nodes are assumed to be malicious, i.e., before transmitting a coded packet they may modify its payload. Nodes receiving corrupted coded packets are prevented from correctly decoding the original chunk. We propose SIEVE, a fully distributed technique to identify malicious nodes.

SIEVE is based on special messages called *checks* that nodes periodically transmit. A check contains the list of nodes identifiers that provided coded packets of a chunk as well as a flag to signal if the chunk has been corrupted. SIEVE operates on top of an otherwise reliable architecture and it is based on the construction of a *factor graph* obtained from the collected checks on which an incremental belief propagation algorithm is run to compute the probability of a node being malicious. Analysis is carried out by detailed simulations using ns-3. We show that SIEVE is very accurate and discuss how nodes speed impacts on its accuracy. We also show SIEVE robustness under several attack scenarios and deceiving actions. The paper has been accepted to [20]

5.4. CCN-TV: a data-centric approach to real-time video services

Participants: Luigi Liquori, Vincenzo Ciancaglini [contact], Riccardo Loti, Giuseppe Piro [Politech Bari], Alfredo Grieco [Politech Bari].

Content Centric Networking is a promising data- centric architecture, based on in-network caching, name-driven routing, and receiver-initiated sessions, which can greatly enhance the way Internet resources are currently used, thus making the support for a broader set of users with increasing traffic demands possible. The CCN vision is, currently, attracting the attention of many researchers across the world, because it has all the potential to become ready to the market, to be gradually deployed in the Internet of today, and to facilitate a graceful transition from a host-centric networking rationale to a more effective data-centric working behavior. At the same time, several issues have to be investigated before CCN can be safely deployed at the Internet scale. They include routing, congestion control, caching operations, name-space planning, and application design. With reference to application-related facets, it is worth to notice that the demand for TV services is growing at an exponential rate over the time, thus requiring a very careful analysis of their performance in CCN architectures. To this end, in the present contribution we deploy a CCN-TV system, able to deliver real-time streaming TV services and we evaluate its performance through a simulation campaign based on real topologies. The paper has been accepted to [19].

5.5. Towards a Trust and Reputation Framework for Social Web Platforms and @-economy

Participants: Thao Nguyen [contact], Bruno Martin [Unice], Luigi Liquori, Karl Hanks.

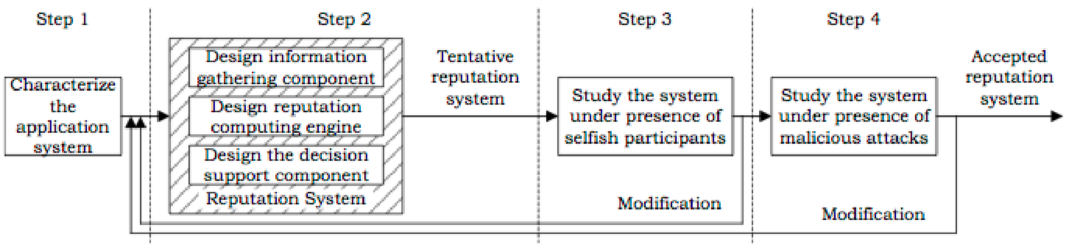


Figure 12. Process of designing a robust trust and reputation system

Trust and reputation systems (TRSs) have recently seen as a vital asset for the safety of online interaction environment. They are present in many practical applications, e.g., e-commerce and social web. A lot of more complicated systems in numerous disciplines also have been studied and proposed in academia. They work as a decision support tool for participants in the system, helping them decide whom to trust and how trustworthy the person is in fulfilling a transaction. They are also an effective mechanism to encourage honesty and cooperation among users, resulting in healthy online markets or communities. The basic idea is to let parties rate each other so that new public knowledge can be created from personal experiences. The greatest challenge in designing a TRS is making it robust against malicious attacks. In this paper, we provide readers an overview on the research topic of TRSs, propose a consistent research agenda in studying and designing a robust TRS, and present an implemented reputation computing engine alongside simulation results, which is our preliminary work to acquire the target of a trust and reputation framework for social web applications.

Information concerning the reputation of individuals has always been spread by word-of-mouth and has been used as an enabler of numerous economic and social activities. Especially now, with the development of technology and, in particular, the Internet, reputation information can be broadcast more easily and faster than ever before. Trust and Reputation Systems (TRSs) have gained the attention of many information and computer scientists since the early 2000s. TRSs have a wide range of applications and are domain specific. The multiple areas where they are applied, include social web platforms, e-commerce, peer-to-peer networks, sensor networks, ad-hoc network routing, and so on. Among these, we are most interested in social web platforms. We observe that trust and reputation is used in many online systems, such as online auction and shopping websites, including eBay, where people buy and sell a broad variety of goods and services, and Amazon, which is a world famous online retailer. Online services with TRSs provide a better safety to their users. A good TRS can also create incentives for good behavior and penalize damaging actions. Markets with the support of TRSs will be healthier, with a variety of prices and quality of service. TRSs are very important for an online community, with respect to the safety of participants, robustness of the network against malicious behavior and for fostering a healthy market.

From a functional point of view, a TRS can be split into three components The first component gathers feedback on participants' past behavior from the transactions that they were involved in. This component includes storing feedback from users after each transaction they take part in. The second component computes reputation scores for participants through a Reputation Computing Engine (RCE), based on the gathered information. The third component processes the reputation scores, implementing appropriate reward and punishment policies if needed, and representing reputation scores in a way which gives as much support as possible to users' decision-making. A TRS can be centralized or distributed. In centralized TRSs, there is a central authority responsible for collecting ratings and computing reputation scores for users. Most of the TRSs currently on the Internet are centralized, for example the feedback system on eBay and customer reviews on Amazon. On the other hand, a distributed TRS has no central authority. Each user has to collect ratings and compute reputation scores for other users himself. Almost all proposed TRSs in the literature are distributed.

Some of the main unwanted behaviors of users that might appear in TRSs are: *free riding* (people are usually not willing to give feedback if they are not given an incentive to do so), *untruthful rating* (users give incorrect feedback either because of malicious intent or because of unintended and uncontrolled variables), *colluding* (a group of users coordinate their behavior to inflate each other's reputation scores or bad-mouth other competitors. Colluding motives are only clear in a specific application), *whitewashing* (a user creates a new identity in the system to replace his old one when the reputation of the old one has gone bad), *milking reputation* (at first, a participant behaves correctly to get a high reputation and then turns bad to make a profit from their high reputation score). The milking reputation behavior is more harmful to social network services and e-commerce than to the others.

This research aims to build on these studies and systematize the process of designing a TRS in general as depicted in Fig. 12 . First, we characterize the application system into which we want to integrate a TRS, and find and identify new elements of information which substitute for traditional signs of trust and reputation in the physical world. Second, based on the characteristics of the application, we find suitable working mechanisms and processes for each component of the TRS. This step should answer the following questions: “What kind of information do we need to collect and how?", “How should the reputation scores be computed using the collected information?", and “How should they be represented and processed to lead users to a correct decision?". To answer the first question, which corresponds to the information gathering component, we should take advantage of information technology to collect the vast amounts of necessary data. An RCE should meet these criteria: *accuracy* for long-term performance (distinguishing a newcomer with unknown quality from a low-quality participant who has stayed in the system for a long time), *weighting* towards recent behavior, *smoothness* (adding any single rating should not change the score significantly), and *robustness* against attacks. Third, we study the tentative design obtained after the second step in the presence of selfish behaviors. During the third step, we can repeatedly return to Step 2 whenever appropriate until the system reaches a desired performance. The fourth step will refine the TRS and make it more robust against malicious attacks. If a modification is made, we should return to Step 2 and check all the conditions in steps 2 and 3 before accepting the modification. The paper has been accepted to [19]

5.6. An Open Logical Framework

Participants: Luigi Liquori [contact], Marina Lenisa [Univ. Udine], Furio Honsell [Univ. Udine], Petar Maksimovic, Ivan Scagnetto [Univ. Udine].

The LFP Framework is an extension of the Harper-Honsell-Plotkin's Edinburgh Logical Framework LF with external predicates, hence the name Open Logical Framework. This is accomplished by defining lock type constructors, which are a sort of "diamond"-modality constructors, releasing their argument under the condition that a possibly external predicate is satisfied on an appropriate typed judgement. Lock types are defined using the standard pattern of constructive type theory, i.e. via introduction, elimination, and equality rules. Using LFP, one can factor out the complexity of encoding specific features of logical systems which would otherwise be awkwardly encoded in LF, e.g. side-conditions in the application of rules in Modal Logics, and sub-structural rules, as in non-commutative Linear Logic. The idea of LFP is that these conditions need only to be specified, while their verification can be delegated to an external proof engine, in the style of the Poincaré Principle or Deduction Modulo. Indeed such paradigms can be adequately formalized in LFP. We investigate and characterize the meta-theoretical properties of the calculus underpinning LFP: strong normalization, confluence, and subject reduction. This latter property holds under the assumption that the predicates are well-behaved, i.e. closed under weakening, permutation, substitution, and reduction in the arguments. Moreover, we provide a canonical presentation of LFP, based on a suitable extension of the notion of $\beta\eta$-long normal form, allowing for smooth formulations of adequacy statements.

LFP is parametric over a potentially unlimited set of (well-behaved) predicates P, which are defined on derivable typing judgements of the form $\Gamma \vdash_\Sigma N : \sigma$, see Fig 13 .

$$\frac{\Gamma \vdash_\Sigma M : \rho \qquad \Gamma \vdash_\Sigma N : \sigma}{\Gamma \vdash_\Sigma \mathcal{L}^{\mathcal{P}}_{N,\sigma}[M] : \mathcal{L}^{\mathcal{P}}_{N,\sigma}[\rho]} \quad (\text{O}\cdot\text{Lock})$$

$$\frac{\Gamma \vdash_\Sigma M : \mathcal{L}^{\mathcal{P}}_{N,\sigma}[\rho] \qquad \Gamma \vdash_\Sigma N : \sigma \qquad \mathcal{P}(\Gamma \vdash_\Sigma N : \sigma)}{\Gamma \vdash_\Sigma \mathcal{U}^{\mathcal{P}}_{N,\sigma}[M] : \rho} \quad (\text{O}\cdot\text{Unlock})$$

Figure 13. Some rule of the Open Logical Framework

The syntax of LFP predicates is not specified, with the main idea being that their truth is to be verified via a call to an external validation tool; one can view this externalization as an oracle call. Thus, LFP allows for the invocation of external "modules" which, in principle, can be executed elsewhere, and whose successful verification can be acknowledged in the system via L-reduction. Pragmatically, lock types allow for the factoring out of the complexity of derivations by delegating the {checking, verification, computation} of such predicates to an external proof engine or tool. The proof terms themselves do not contain explicit evidence for external predicates, but just record that a verification {has to be (lock), has been successfully (unlock)} carried out. In this manner, we combine the reliability of formal proof systems based on constructive type theory with the efficiency of other computer tools, in the style of the Poincaré Principle. In this paper, we develop the meta-theory of LFP. Strong normalization and confluence are proven without any additional assumptions on predicates. For subject reduction, we require the predicates to be well-behaved, i.e. closed under weakening, permutation, substitution, and $\beta\mathcal{L}$-reduction in the arguments. LFP is decidable, if the external predicates are decidable. We also provide a canonical presentation of LFP, based on a suitable extension of the notion of $\beta\eta$-long normal form. This allows for simple proofs of adequacy of the encodings. In particular, we encode in LFP the call-by-value λ-calculus and discuss a possible extension which supports the design-by-contract paradigm. We provide smooth encodings of side conditions in the rules of Modal Logics, both in Hilbert and Natural Deduction styles. We also encode sub-structural logics, i.e. non-commutative Linear Logic. We also illustrate how LFP can naturally support program correctness systems and Hoare-like logics. In our encodings, we utilize a library of *external predicates*. As far as expressiveness is concerned, LFP is a stepping stone towards a general theory of shallow vs deep encodings, with our encodings being shallow by definition. Clearly, by Church's thesis, all external decidable predicates in LFP can be encoded, possibly with very deep encodings, in standard LF. It would be interesting to state in a precise categorical setting the relationship between such deep internal encodings and the encodings in LFP. LFP can also be viewed as a neat methodology for separating the logical-deductive contents from, on one hand, the verification of structural and syntactical properties, which are often needlessly cumbersome but ultimately computable, or, on the other hand, from more general means of validation.

MADYNES Project-Team

6. New Results

6.1. Android Security

Participants: Olivier Festor, Abdelkader Lahmadi [contact].

Android-based devices include smartphones and tablets that are now widely adopted by users because they offer a huge set of services via a wide range of access networks (WiFi, GPRS/EDGE, 3G/4G). Android provides the core platform for developing and running applications. Those applications are available to the users over numerous online marketplaces. These applications are posted by developers, with little or no review process in place, leaving the market self-regulated by users. This policy generates a side-effect where users are becoming targets of different malicious applications which the goal is to steal their private information, collect all kind of sensitive data via sensors or abusing granted permissions to make surtaxed calls or messages. To address this security issue, monitoring the behaviour of running applications is a key technique enabling the identification of malicious activities.

During 2012, we have designed and developed a monitoring framework integrating observed network and system activities of a running application. We have developed an embedded NetFlow probe running on android devices to export observed network flow records observed to a collection point for their processing. Our embedded probe includes a new set of IPFIX information elements that we have designed [36] to encapsulate location information within exported flows using the IPFIX protocol.

We have also developed an embedded logging probe that exports available system logs to a collection point. The logs are then centrally processed and correlated with observed network flow records to extract an accurate behavior of an application including its network and in-device activities.

Our monitoring framework is different from available proposed solutions since we build a dynamic model to infer the running behavior of an Android application. This technique allows us to identify patched applications where a malicious activity has been added, cloned applications where the observed behavior is different from the expected behavior and privacy leaks where an application is contacting unexpected services.

6.2. Sensor networks monitoring

Participants: Alexandre Boeglin, Laurent Ciarletta, Olivier Festor, Abdelkader Lahmadi [contact], Emmanuel Nataf, Bilel Saadallah.

Low Power and Lossy Networks (LLNs) are made of interconnected wireless devices with limited resources in terms of energy, computing and communication. The communication channels are low-bandwidth, high loss rate and volatile wireless links subject to failure over time. They are dynamic and the connectivity is limited and fluctuant over time. Each node may loss frequently its connectivity with its neighborhood nodes. In addition, link layer frames have high constrains on their size and throughput is limited. These networks are used for many different applications including industrial automation, smart metering, environmental monitoring, homeland security, weather and climate analysis and prediction. The main issue in those networks is optimal operation combined with strong energy preservation. Monitoring, i.e the process of measuring sampled properties of nodes and links in a network, is a key technique in operational LLNs where devices need to be constantly or temporally monitored to assure their functioning and detect relevant problems which will result in an alarm being for- warded to the enterprise network for analysis and remediation.

During the year 2012, we developed novel approaches for the monitoring of LLNs. We developed and designed a novel algorithm and a supporting framework [18] that improves a poller-pollee monitoring architecture. We empower the poller-pollee placement decision process and operation by exploiting available routing data to monitor nodes status. In addition, monitoring data is efficiently embedded in any messages flowing through the network, drastically reducing monitoring overhead. Our approach is validated through both simulation, implementation and deployment on a 6LoWPAN-enabled network. Both simulations and large-scale testbed experiments assess the efficiency of our monitoring scheme. Results also demonstrate that our approach is less aggressive and less resource consuming than its competitors.

We developed a first fully operational CCNx stack [40] on a wireless sensor network. We implemented CCNx as a native C experimental extension of Contiki, an operating system dedicated to Internet of Things applications. Our extension [33] is based on the reference implementation of CCNx modified to run as a network driver on top of different available MAC protocols implementations in Contiki. Our goal is to design a monitoring and configuration framework that benefits from the content-centric approach to efficiently collect desired management content and apply in-network processing functions for nodes configuration and monitoring. This includes extending naming schema with monitoring oriented processing functions, optimizing data interests to minimize the communication overhead.

6.3. Management and monitoring of P2P networks

Participants: Isabelle Chrisment [contact], Olivier Festor, Juan Pablo Timpanaro.

In 2012, we have addressed operation, monitoring and security issues on several P2P target networks: KAD, BitTorrent and I2P.

Several large scale P2P networks operating on the Internet are based on a Distributed Hash Table. These networks offer valuable services, but they all suffer from a critical issue allowing malicious nodes to be inserted in specific places on the DHT for undesirable purposes (monitoring, distributed denial of service, pollution, etc.). While several attacks and attack scenarios have been documented, few studies have measured the actual deployment of such attacks and none of the documented countermeasures have been tested for compatibility with an already deployed network. In our work, we focus on the KAD network. Based on large scale monitoring campaigns, we demonstrated that the world-wide deployed KAD network suffers large number of suspicious insertions around shared contents and we quantify them. To cope with these peers, we proposed a new efficient protection algorithm based on analyzing the distribution of the peers ID found around an entry after a DHT lookup [3]. The evaluation of our solution showed that it detects the most efficient configurations of inserted peers with a very small false-negative rate, and that the countermeasures successfully filter almost all the suspicious peers. We demonstrate the direct applicability of our approach by implementing and testing our solution in real P2P networks

BitTorrent is a fast, popular, P2P filesharing application focused on fast propagation of content. Its trackerless approach uses a DHT based on Kademlia to search for sources when the hash of the metadata of the content to transfer is known. On the other hand, the eMule network uses the old ED2K protocol for filesharing including a system of priorized queues, but indexation is done through a solid Kademlia based DHT, named Kad. The Kad DHT stands for a search engine, wich provides an extra level to map keywords to file identifiers. We have designed an hybrid approach, compatible with both P2P file-sharing networks, which has the Kad advantages on indexation and the BitTorrent throughput for transfer while maintaining backward compatibility with both of these networks [42]. To validate our proposal we developed a prototype which supports content indexation provided by the Kad network and is able to transfer files using the BitTorrent protocol. Using this prototype, we measured the propagation of new content in clusters of aMule clients, BitTorrent clients, hybrid clients, and a mix of them.

In parallel, we continued our research about being anonymous when downloading from BitTorrent. Anonymous communications have been gaining more and more interest from Internet users as privacy and anonymity problems have emerged. Among anonymous enabled services, anonymous file-sharing is one of the most active one and is increasingly growing. Large scale monitoring on these systems allows us to grasp how they behave, which type of data is shared among users, the overall behavior in the system.

We presented the first monitoring study aiming to characterize the usage of the I2P network, a low-latency anonymous network based on garlic routing [23]. We characterized the file-sharing environment within I2P, and evaluated if this monitoring affects the anonymity provided by the network. We showed that most activities within the network are file-sharing oriented, along with anonymous web-hosting. We assessed the wide geographical location of nodes and network popularity. We also demonstrated that group-based profiling is feasible on this particular network [22].

Dedicated anonymous networks such as Freenet and I2P allow anonymous file-sharing among users. However, one major problem with anonymous file-sharing networks is that the available content is highly reduced, mostly with outdated files, and non-anonymous networks, such as the BitTorrent network, are still the major source of content. We showed that in a 30-days period, 21648 new torrents were introduced in the BitTorrent community, whilst only 236 were introduced in the anonymous I2P network, for four different categories of content. Therefore, how can a user of these anonymous networks access this varied and non-anonymous content without compromising its anonymity? In [24], we improved content availability in an anonymous environment by proposing the first internetwork model allowing anonymous users to access and share content in large public communities while remaining anonymous. We showed that our approach can efficiently interconnect I2P users and public BitTorrent swarms without affecting their anonymity nor their performance. Our model is fully implemented and freely usable.

6.4. Configuration security automation

Participants: Rémi Badonnel [contact], Martin Barrere, Olivier Festor.

The main research challenge addressed in this work is focused on enabling configuration security automation in autonomic networks and services. In particular our objective is to increase vulnerability awareness in the autonomic management plane in order to prevent configuration vulnerabilities. The continuous growth of networking significantly increases the complexity of management. It requires autonomic networks and services that are capable of taking in charge their own management by optimizing their parameters, adapting their configurations and ensuring their protection against security attacks. However, the operations and changes executed during these self-management activities may generate vulnerable configurations. A first part of our work in the year 2012 has been dedicated to the assessment of distributed vulnerabilities and to the elaboration of a collaborative management strategy for supporting their remediation. A configuration vulnerability is not necessarily local but can also be spread over several devices in the autonomic network. We have showed in [8] how such distributed vulnerabilities can be mathematically formalized and described in a machine readable manner, through the specification of the DOVAL (Distributed OVAL) language on top of OVAL (Open Vulnerability and Assessment Language). We have designed and evaluated a dedicated framework for exploiting these vulnerability descriptions, collecting device configurations and detecting distributed vulnerabilities using specific aggregation techniques. Once a vulnerability is identified in the autonomic network, several remediation actions can potentially be performed by the autonomic network over devices. For that purpose, we have introduced an XCCDF-based specification for expressing alternative treatments related to a distributed vulnerability. We have also proposed a collaborative scheme for selecting one of these treatments depending on the current context (device capabilities and willingness to participate) [6]. A second part of our work has focused on the extension of our solution to other environments. In particular we have worked on the integration of our vulnerability assessment strategy over the Android plateform [9]. We have put forward a mathematical model as well as an optimized method that provides solid foundations for this context. By maintaining low-consumption services monitoring the system, the proposed approach minimizes heavy task executions by only triggering assessment activities when configuration changes are detected or new vulnerability definitions are available. In light of this, we have developed a prototype that efficiently performs self-assessment activities, and also introduces dedicated web services for collecting OVAL descriptions and storing assessment results. We have performed an analytical evaluation of the proposed model as well as an extensive set of technical experiments that shows the feasibility of our solution. We are currently working on the issue of past hidden vulnerable states. A network compromised in the past by an unknown vulnerability at that moment may still constitute a potential security threat in the present. Accordingly, past unknown system exposures are required to be taken into account. We are therefore investigating a novel strategy for identifying also such past hidden vulnerable configurations and increasing the overall security [9].

6.5. Cache Management in CCN

Participants: Thomas Silverston [contact], César Bernardini, Olivier Festor.

The Internet is currently mostly used for accessing content. Indeed, ranging from P2P file sharing to current video streaming services such as Youtube, it is expected that content will count for approximately 86% of the global consumer traffic by 2016.

While the Internet was designed for -and still focuses on- host-to-host communication (IP), users are only interested in actual content rather than source location. Hence, new Information-Centric Networking architectures (ICN) such as CCN, NetInf, Pursuit have been proposed giving high priority to efficient content distribution at large scale. Among all these new architectures, Content Centric Networking (CCN) has attracted considerable attention from the research community [2].

CCN is a network architecture based on named data where a packet address names content, not location. The notion of host as defined into IP does not exist anymore. In CCN, the content is not retrieved from a dedicated server, as it is the case for the current Internet. The premise is that content delivery can be enhanced by including per-node-caching as content traverses the network. Content is therefore replicated and located at different points of the network, increasing availability for incoming requests.

As content is cached along the path, it is crucial to investigate the caching strategy for CCN Networks and to propose new schemes adapted to CCN. We therefore designed *Most Popular Content* (MPC), a new caching strategy for CCN network [10].

Instead of storing all the content at every nodes on the path, MPC strategy caches only popular content. With MPC, each nodes count all the requests for a content and when it has been requested a large amount of time, the content will be cached at each node along the path. Otherwise, the content is not popular; it is transmitted but it is not cached into the network.

We implemented MPC into the ccnSim simulator and evaluate it through extensive simulations.

Our results demonstrate that using MPC strategy allow to achieve a higher Cache Hit in CCN networks and still reduces drastically the number of replicas. By caching only popular content, MPC helps at reducing the cache load at each node and the network resource consumption.

We expect that our strategy could serve as a base for studying name-based routing protocols. Being a suggestion based mechanism, it is feasible to adapt it to manage content among nodes, to predict popularity and to route content to destination. In addition, we are currently investigating the social relationship between users to improve our caching strategy for CCN networks.

6.6. QoS in Wireless Sensor Networks

Participants: François Despaux, Abdelkader Lahmadi, Bilel Nefzi, Hugo Cruz-Sanchez, Ye-Qiong Song [contact].

WSN research focus has progressively been moved from the energy issue to the QoS issue. Typical example is the MAC protocol design, which cares about not only low duty-cycle, but also high throughput with self-adaptation to dynamic traffic changes [21]. Our research on WSN QoS is thoroughly organized in three topics:

- MAC protocol design for both QoS and energy efficiency

[2] http://www.ccnx.org

The main result that we obtained in 2012 is a new hybrid CSMA/TDMA MAC protocol, called Queue-MAC, that dynamically adapts the duty-cycle according to the current network traffic. The queue length of nodes is used as the network traffic indicator. When the traffic increases, the active CSMA period is accordingly extended by adding dynamic TDMA slots, allowing thus to efficiently handle burst traffic under QoS constraints. This protocol is implemented on the STM32W108 SOC chips and compared with both a fixed duty-cycle reference protocol and an optimized IEEE802.15.4 MAC protocol. Through extensive experimental measurements, we showed that our queue-length aware hybrid CSMA/TDMA MAC protocol largely outperforms the compared protocols. The proposed protocol can be easily implemented through slight adaptation of the IEEE802.15.4 standard [25].

Many industrial WSN are based on IEEE802.15.4 standard. One of the critical issues is the scheduling of neighboring coordinators beacons. In [20], we presented TBoPS, a novel technique for scheduling beacons in the cluster-tree topology. TBoPS uses a dedicated period called beacon only period (BOP) to schedule beacons at the beginning of IEEE 802.15.4 superframe. The advantage of TBoPS is that every beacon-enabled node distributively selects a beacon schedule during association phase.

- QoS routing

  For supporting different QoS requirements, routing in WSN must simultaneously consider several criteria (e.g., minimizing energy consumption, hop counts or delay, packet loss probability, etc.). When multiple routing metrics are considered, the problem becomes a multi-constrained optimal path problem (MCOP), which is known as NP-complete. In practice, the complexity of the existing routing algorithms is too high to be implemented on the low cost and power constrained sensor nodes. Recently, Operator calculus (OC) has been developed by Schott and Staples with whom we collaborate. OC can be applied to solving MCOP problem with much lower complexity and can deal with dynamic topology changes (which is the case in duty-cycled WSN). The OC approach has been successfully applied to a concrete routing problem [13]. Its implementation over Contiki on TelosB motes has also been achieved, confirming thus its great potential for developing new QoS routing protocols for WSN.

- End-to-end performance in multi-hop networks

  Probabilistic end-to-end performance guarantee may be required when dealing with real-time applications. For instance, in our ANR QUASIMODO project, we considered an intrusion detection and tracking scenario and analyzed the application requirements with respect to the network QoS. Assuming the use of the extended Kalman filter based tracking technique, we derived the tradeoff relationship between the tracking precision and the delay (from the target position and speed sampling to mobile nodes moving to cover the estimated next step area). In [5] we proposed a novel coordinative moving algorithm for autonomous mobile sensor networks to guarantee that the target can be detected in each observed step while minimizing the amount of moving sensors (so saving energy). In such kind of application context, we aim to provide methods for both network resource allocation and estimating the end-to-end delay in multi-hop WSN. Assuming IEEE802.15.4 WSN with cluster-tree routing, in [16] we addressed the problem of allocating and reconfiguring the available bandwidth using an Admission Control Manager that guarantees that the nodes respect their probabilistic bandwidth assignment when generating data traffic. It has been shown by simulation that using the proposed method, one can obtain desired probabilistic guarantee in both bandwidth and energy efficiency.

  In a more general context of meshed networks, we present an empirical support of an analytical approach, which employs a frequency domain analysis for estimating end-to-end delay in multi-hop networks. The proposed analytical results of the end-to-end delay distribution are validated through simulation and compared with queuing theory based analysis. Our results demonstrate that an analytical prediction schema is insufficient to provide an adequate estimation of the end-to-end delay distribution function, but it requires to be combined with simulation methods for detailed link and node latency distribution [15].

6.7. Energy in Wireless Sensor Networks

Participants: Emmanuel Nataf [contact], Patrick-Olivier Kamgueu.

The energy sources of sensors in a wireless network rely mainly on batteries and are very limited in their capacity. Several research efforts are focalized on trying to limit the energy consumption in such networks. This is particularly the case in protocol design. Indeed, the communication consumes a large majority of the available energy. To be realistic and efficient, all proposed approaches need to know the energy available at any time in the systems. Unfortunately, most sensors do not provide such information because it requires additional built-in hardware that would drastically increase their cost. Over the last decade very accurate physical battery models that encompass consumption and recovery have been designed. The complexity of these models is however too high to be implemented inside simple sensors. Recent research results have shown that this integration could be possible if some approximations are integrated in the models.

We have worked on integrating such an approcimated model in the sensor operating system. This work allows the simulation of such sensors and the deployement on real devices that will be aware of their remaining energy level without requiring any additionnal costly equipment. A first implementation on simulation tool has given very promising results; sensors can access their energy level and take decision based on this estimate. Firstly, we have studied energy consumption of a sensors network collecting and routing data toward a single destination. Energy cost of the network deployment has been computed and so the network life as a whole. An other result of our work is the comparison of several common link layer access protocols and several data rate transmits [31].

6.8. Online Risk Management

Participants: Rémi Badonnel [contact], Oussema Dabbebi, Olivier Festor.

Telephony over IP has known a large scale deployment and has been supported by the standardization of dedicated signaling protocols. This service is however exposed to multiple attacks due to a lower confinement in comparison to traditional PSTN networks. While a large variety of methods and techniques has been proposed for protecting VoIP networks, their activation may seriously impact on the quality of such a critical service. Risk management provides new opportunities for addressing this challenge. In particular our work aims at performing online risk management for VoIP networks and services. The objective is to dynamically adapt the service exposure with respect to the threat potentiality, while maintaining a low security overhead. In the year 2012, we have pursued our work on online risk management and applied it to more distributed configurations. In that context we have defined in [14] an exposure control solution for P2PSIP networks where the registration and location servers are implemented by a distributed hash table. After having analyzed different attack scenarios, we have designed the underlying risk management architecture and modelled several dedicated countermeasures. We have evaluated the performance and scalability of our approach through extensive experiments performed with the OMNET++ simulator. We have also proposed a trust-based solution for addressing residual attacks in the RELOAD framework. This latter, complementary to our risk management approach, is a peer-to-peer signalling overlay using a central certificate enrolment server and supporting P2PSIP infrastructures. Self-signed certificates can also be used in closed networks, and connections amongst nodes can be secured using an encryption protocol such as TLS. While the RELOAD framework permits to reduce the exposure to threats, P2PSIP networks are still exposed to residual attacks related to the routing and storage activities. For instance, it is trivial for a malicious node to refuse to give the stored information, or to send false routing messages in the network. We have showed how trust mechanisms can be exploited to counter these attacks in an efficient manner. Our work on online risk management has also focused on VoIP services in the Cloud [30]. The integration of IP telephony in this environment permits the delivery and access of new resources and constitutes an important factor for its scalability. While the Cloud has recently served as a basis for security attacks targeting IP telephony, such as SIP brute force attacks from the Amazon EC2 Cloud infrastructure, we consider that it also provides new possibilities for supporting the security of this service. We have analyzed the applicability of our online risk management approach in the Cloud, and evaluated to what extent security countermeasures may be outsourced as a service. We have mathematically defined a dedicated modelling and detailed different treatment strategies for applying countermeasures in the Cloud. Finally, we have quantified the benefits and costs of these strategies based on a set of experimental results.

6.9. Pervasive Computing

Participants: Laurent Ciarletta [contact], Olivier Festor, Ye-Qiong Song, Adrien Guenard, Yannick Presse.

Vincent Chevrier(MAIA Team), Thomas Navarrette Gutierrez (MAIA Team) and Priyadrsi Nanda (Universit of Technology, Sydney) did contribute to part of this activity.

In Pervasive or Ubiquitous Computing, a growing number of communicating/computing devices are collaborating to provide users with enhanced and ubiquitous services in a seamless way. In a related field, Cyber Physical Systems also are technological systems that have to be considered within a physical world and its contraints. They are complex systems where several inter-related phenomena have to be considered. In order to be studied, modeled and evaluated, we propose the use of co-simulation and multimodeling. to be Madynes is focusing on the networking aspects of such systems. We cooperate with toher the Maia team to be able to encompass issues and research questions that combine both networking and cognitive aspects.

Pervasive Computing is about interconnected and situated computing resources providing us(ers) with contextual services. These systems, embedded in the fabric of our daily lives, are complex: numerous interconnected and heterogeneous entities are exhibiting a global behavior impossible to forecast by merely observing individual properties. Firstly, users physical interactions and behaviors have to be considered. They are influenced and influence the environment. Secondly, the potential multiplicity and heterogeneity of devices, services, communication protocols, and the constant mobility and reorganization also need to be addressed. Our research on this field is going towards both closing the loop between humans and systems, physical and computing systems, and taming the complexity, using multi-modeling (to combine the best of each domain specific model) and co-simulation (to design, develop and evaluate) as part of a global conceptual and practical toolbox. We apply this work on UAvs and energy-constrained / location aware services.

In 2012 we worked on the following research topics :

- Continuing the work on multi-modeling and co-simulation, we've participated with the MAIA team on the development of an architecture for the control of complex systems based on multi-agent simulation [32], [2], and a CPS co-simulation (next item), and continue working on the AA4MM framework (Agents and artefacts for Multiple heterogeneous Models).
- In Cyper Physical Systems, we have lead the design and implementation of the Aetournos (Airborne Embedded auTonomOUs Robust Network of Objects and Sensors) platform at Loria. The idea of AETOURNOS is to build a platform which can be at the same time a demonstrator of scientific realizations and an evaluation environment for research works of various teams of our laboratory. It is also its own research domain : building a completely autonomous and robust flock of collaborating UAVs.

  In Madynes, we focus on the CPS and their networks and applications. Those systems consist of numerous autonomous elements in sharp interaction which functioning require a tight coupling between software implementations and technical devices. The collective movements of a flock of flying communicating robots / UAVs, evolving in potentially perturbed environment constitute a good example of such a system. Indeed, if we look at the level of each of the elements playing a role into this system, a certain number of challenges and scientific questions can be studied: respect of real-time constraints of calculations for every autonomous UAV and for the communication between the robots, conception of individual, embedded, distributed or global management systems, development of self-adaptative mechanisms, conception of algorithms of collective movement etc... Furthermore, the answers to each of these questions have to finally contribute to the global functioning of the system. Applying co-simulation technique we plan to develop a hybrid "network-aware flocking behavior" / "behavior aware routing protocol". The platform is composed of several high-grade research UAVs (Pelican quadcopters and Firefly hexacopters) and lighter models (AR.Drone quacopters). We have provided a working set of tools : multi-simulation behavior / network / physics and generic software development using ROS (Robot Operating System). The UAVs carry a set of sensor for location awareness, their own computing capabilities and several wireless networks.

This work is discribed in a position paper where a first implementation of a formation flight is detailed ([11]).

- Energy-constraint geolocalization, addressing, routing and management of wireless devices: a research collaboration with Fireflies RTLS was started in March 2009 and has ended in 2012. The initial work has been extended in a joint work with the former TRIO Team and a visiting professor from the University of Technology of Sydney. Its focus has been shiffted towards novel adressing and routing scheme minimizing a global energy-cost function in a wireless sensor network location systems [28]. We are proposing a global configuration tool for this matter in regards with given constraints (number of nodes, topology, QoS).

In 2013, we will continue working on the hybrid protocols and on the UAV platform, and apply our co-simulation work to Smart Grids.

MAESTRO Project-Team

5. New Results

5.1. Network Science

Participants: Eitan Altman, Konstantin Avrachenkov, Mahmoud El Chamie, Philippe Nain, Giovanni Neglia, Marina Sokol.

5.1.1. Epidemic models of propagation of content

E. Altman and P. Nain have studied in [96] in collaboration with A. Shwartz (Technion, Israel) and Y. Xu (Univ. Avignon/LIA) the efficiency of the existing methods for reducing availability of non-authorized copyrighted content for free download on the Internet. To model the propagation of the content, they used both branching processes as well as several epidemic models. One of the important finding is that the greatest impact of measures against unauthorized download is obtained whenever some parameter that describes the virality of the content is close to some critical value (which is computed in this work).

5.1.2. Control and game models for malware attack

In collaboration with M. H. R. Khouzani (Ohio State Univ., USA) and S. Sarkar (Univ. of Pennsylvania, USA), E. Altman has used in [31],[33], [32], optimal control theory to study malware attack in networks. The structure of optimal policies is obtained by using the Pontryagin maximum principle. In the first two references, optimal defense policies are studies in the goal of protecting the network. In the third work, the worst case behavior of the attack is identified using control theory. The authors then study in [34] the combined problem of identifying the defensive control that achieves the best performance under the worst possible malware attack. This is done through a zero-sum game context.

5.1.3. Time random walks on time varying graphs

In collaboration with D. Figueiredo (Federal Univ. of Rio de Janeiro, Brazil), B. Ribeiro and D. Towsley (both from the Univ. of Massachusetts at Amherst, USA), P. Nain has studied the behavior of a continuous time random walk (CTRW) on a stationary and ergodic time varying dynamic graph [57]. Conditions have been established under which the CTRW is a stationary and ergodic process. In general, the stationary distribution of the walker depends on the walker rate and is difficult to characterize. However, the stationary distribution has been characterized in the following cases: i) the walker rate is significantly larger or smaller than the rate in which the graph changes (time-scale separation), ii) the walker rate is proportional to the degree of the node that it resides on (coupled dynamics), and iii) the degrees of nodes belonging to the same connected component are identical (structural constraints). Examples are provided that illustrate these theoretical findings.

5.1.4. Quick detection of central nodes

In [50] K. Avrachenkov and M. Sokol, together with N. Litvak (Twente Univ., The Netherlands) and D. Towsley (Univ. of Massachusetts at Amherst, USA) propose a random walk based method to quickly find top k lists of nodes with the largest degrees in large complex networks. The authors show theoretically and by numerical experiments that for large networks the random walk method finds good quality top lists of nodes with high probability and with computational savings of orders of magnitude. They also propose stopping criteria for the random walk method which requires very little knowledge about the structure of the network.

5.1.5. Graph-based semi-supervised learning

In [48] K. Avrachenkov and M. Sokol, together with P. Gonçalves (INRIA project-team RESO) and A. Mishenin (St. Petersburg State Univ., Russia) develop a generalized optimization framework for graph-based semi-supervised learning. The framework gives as particular cases the Standard Laplacian, Normalized Laplacian and PageRank based semi-supervised learning methods. The authors provide new probabilistic interpretation based on random walks and characterize the limiting behaviour of the methods. The random walk based interpretation allows one to explain differences between the performances of methods with different smoothing kernels. It appears that the PageRank based method is robust with respect to the choice of the regularization parameter and the labelled data. The theoretical results are illustrated with two realistic datasets, characterizing different challenges: "Les Misérables" characters social network and Wikipedia hyper-link graph. It appears that the PageRank based method can classify the Wikipedia articles with very good precision and perfect recall employing only the information about the hyper-text links.

In [47] K. Avrachenkov and M. Sokol, together with P. Gonçalves (INRIA project-team RESO) and A. Legout (INRIA project-team PLANETE) apply the theoretical results of [48] to classification of content and users in BitTorrent. The general intuition behind the application of the graph based semi-supervised learning methods is that the users with similar interests download similar contents. PageRank based semi-supervised learning method was chosen as it scales well with very large volumes of data. The authors provide recommendations for the choice of parameters in the PageRank based semi-supervised learning method, and show, in particular, that it is advantageous to choose labelled points with large PageRank score.

5.1.6. Optimal weight selection in average consensus protocols

In average consensus protocols, nodes in a network perform an iterative weighted average of their estimates and those of their neighbors. The protocol converges to the average of initial estimates of all nodes found in the network. The speed of convergence of average consensus protocols depends on the weights selected on links (to neighbors). In [92] K. Avrachenkov, M. El Chamie and G. Neglia address how to select the weights in a given network in order to have a fast speed of convergence for these protocols. They approximate the problem of optimal weight selection by the minimization of the Schatten p-norm of a matrix with some constraints related to the connectivity of the underlying network. They then provide a totally distributed gradient method to solve the Schatten norm optimization problem. By tuning the parameter p in the proposed minimization, it is possible to simply trade-off the quality of the solution (i.e. the speed of convergence) for communication/computation requirements (in terms of number of messages exchanged and volume of data processed). Simulation results on random graphs and on real networks show that this approach provides very good performance already for values of p that only needs limited information exchange. The weight optimization iterative procedure can also run in parallel with the consensus protocol and form a joint consensus–optimization procedure.

5.1.7. Reducing communication overhead of average consensus protocols

The average consensus protocol converges only asymptotically to consensus and implementing a termination algorithm is challenging when nodes are not aware of some global information (e.g. the diameter of the network or the total number of nodes). In [93] K. Avrachenkov, M. El Chamie and G. Neglia propose a totally distributed algorithm for average consensus where nodes send more messages when they have large differences in their estimates, and reduce their message sending rate when the consensus is almost reached. The convergence of the system is guaranteed to be within a predefined margin from the true average and the algorithm gives a trade-off between the precision of consensus and the number of messages send in the network. The proposed algorithm is robust against nodes changing their estimates and can also be applied in dynamic networks with faulty links.

5.2. Wireless Networks

Participants: Eitan Altman, Philippe Nain, Giovanni Neglia.

5.2.1. Estimation of population sizes in sensor networks

We have been working on several problems related to the estimation of population sizes. In collaboration with D. Kumar (IBM Research Center, Hawthorne, USA) and T. Başar (Univ. of Illinois at Urbana-Champaign, USA), E. Altman develops in [73] a Wiener filter that allows to estimate the number of sensors that cover the space at some selected points. The authors take advantage of spatial correlations between the number of sensors covering different points in order to derive the filter. We note that causality is not an issue in space, in contrast to filtering at different points in time.

In collaboration with A. Ali, T. Chahed and M. K. Panda (Telecom SudParis, France), D. Fiems (Gent Univ., Belgium), and L. Sassatelli (I3S, Univ. Nice Sophia Antipolis - CNRS, France), E. Altman has used in [37] Kalman filtering theory in order to estimate the number of mobiles in a delay tolerant ad-hoc network which have a copy of a broadcasted message.

5.2.2. Cellular networks: Small cells

Analysing performance measures of cellular systems combines tools from queueing theory and stochastic processes, on one hand, and geometric considerations on the other hand. In [72], V. Kavitha (Mymo Wireless, Bangalore, India), S. Ramanath (Lekha Wireless Solutions, Bangalore, India), and E. Altman compute the time it takes to transmit a file taking into account the channel conditions which vary due to mobility of terminals. Mobility considerations play a key role in small cells since handover may occur way before the transmission of the file ends.

5.2.3. Multi scale fairness concepts for resource allocation in wireless networks

In many applications that require resources, one needs these resources within some given deadline. These impose constraints when attempting to allocate resources fairly. In [14], E. Altman, K. Avratchenkov and S. Ramanath have extended the α fairness concept by Mo and Walrand so as to include time constraints. They study the question of how to compute such constrained fair allocation, and derive some asymptotic properties of constrained fair assignment.

5.2.4. Self organization in cellular communications

Self organization is an approach to design networks so as to allow them to configure in an automatic way. This allows to reduce the complexity in systems containing thousands of mobiles and a huge number of small cells. In cellular networks, self organization can be used for deciding on time or frequency reuse according to the interference in these time and frequency slots from other cells. The impact of self organization on communications are derived in [55] and [21] by R. Combes, and Z. Altman (Orange Labs, Issy les Moulineaux), in collaboration with E. Altman.

5.2.5. Streaming over wireless

In [75], E. Altman and M. Haddad study in collaboration with T. Jiménez and R. El-Azouzi (Univ. Avignon/LIA) and S.-E. Elayoubi (Orange Labs, Issy les Moulineaux) streaming service over cellular networks. The purpose is to obtain the exact distribution of the number of buffer starvations within a sequence of N consecutive packet arrivals. This is then applied to optimize the quality of experience (QoE) of media streaming service over cellular networks by exploiting the tradeoff between the start-up delay and the starvation.

5.2.6. Wireless network security

The operation of a wireless network relies extensively on exchanging messages over a universally known channel, referred to as the control channel. The network performance can be severely degraded if a jammer launches a denial-of-service (DoS) attack on such a channel.

In [94], P. Nain, M. Krunz, H. Rahbari and M. J. Abdel Rahman (all three from Univ. of Arizona, USA) design frequency hopping (FH) algorithms that mitigate DoS attacks on the control channel of an asynchronous ad hoc network. More specifically, three FH algorithms (called NUDoS, KMDoS, and NCMDoS) are developed for establishing unicast (NUDoS) and multicast (KMDoS and NCMDoS) communications in the presence of multiple jammers. KMDoS and NCMDoS provide different tradeoffs between speed and robustness to node compromise. These algorithms are fully distributed, do not incur any additional message exchange overhead, and can work in the absence of node synchronization. Furthermore, KMDoS and NCMDoS have the attractive feature of maintaining the multicast group consistency. NUDoS exploits the grid quorum system, whereas KMDoS and NCMDoS use the uniform k-arbiter and the Chinese remainder theorem (CRT) quorum systems, respectively. Extensive simulations are used to evaluate these algorithms.

5.3. Network engineering games

Participants: Eitan Altman, Konstantin Avrachenkov, Ilaria Brunetti, Richard Combes, Julien Gaillard, Majed Haddad, Manjesh Kumar Hanawal, Alexandre Reiffers.

5.3.1. Fairness

Anti-trust laws have been introduced by many countries in the last century. This is due to the perception that free competition is better for society. This motivated H. Kameda (Univ. Tsukuba, Japan), C. Touati and A. Legrand (MESCAL, INRIA - CNRS) in cooperation with E. Altman, to define in [28] a fairness concept related to the outcome of competition, which is the Nash equilibrium concept.

5.3.2. Association problem

In [70], E. Altman and M. Haddad study in collaboration with C. Hasan and J.-M. Gorce (SOCRATE, INRIA - INSA) games related to the association problem of mobiles to an access point. It consists of deciding to which access point to connect. Here the choice is between two access points or more, where the access decisions may depend on the number of mobiles connected to each one of the access points. New results were obtained using elementary tools in congestion and crowding games.

5.3.3. Association and placement

The location of a base station has an impact on the throughuput of arriving mobiles that decide to connect to it. Given a cooperative behavior among base stations, E. Altman derives in [54] in collaboration with A. Coluccia (Univ. Salento, Italy) the equilibrium association policy and maximizes its performance by a suitable cooperative positioning of the base stations. The non-cooperative related model was studied in [16] by E. Altman, in collaboration with A. Kumar, C. Singh and R. Sundaresan (all three from IISc, Bangalore, India).

5.3.4. Power control with energy state

In [42] and [64], E. Altman, M. Haddad, J. Gaillard study with D. Fiems (Gent Univ., Belgium) a power control game over a collision channel. Each player has an energy state. When choosing a higher transmission power, the chance of a successful transmission (in the presence of other interference) increases at the cost of a larger decrease in the energy state of the battery. This dynamic game is studied when restricting to simple non-dynamic strategies that consist of choosing a given power level that is maintained during the lifetime of the battery. Surprising paradoxes were identified in the proposed Hawk and Dove game.

5.3.5. Routing games

In [65], M. Haddad, E. Altman and J. Galliard study in collaboration with D. Fiems (Gent Univ., Belgium) a sequential dynamic routing game on a line, where the decision of a user is spatio-temporal control. Each user ships its demand over time on a shared resource. Explicit expressions of the equilibrium of such systems are presented and compared to the global optimum case. The basic idea is taken from a previous paper on this subject by M. K. Hanawal (also with Univ. Avignon/LIA) and E. Altman, in collaboration with R. El-Azouzi (Univ. Avignon/LIA) and B. Prabhu (CNRS - LAAS), who show in [67] that one may transform the time dimension into a spatial component and thus obtain an equivalent standard routing game (where time plays no role) with infinitely many nodes.

5.3.6. Bayesian games in networking

We have considered several problems in networks in which decision makers have asymmetrical information. One of these is how one agent may benefit from revealing part of his information? We considered two types of hierarchical scenarios. In the first, we assume that an agent signals some information to another agent who then chooses an action based on that signal. This action determines the utility of both agents. In the second scenario, a player takes an action (such as pricing) and then the second player reacts to it. Both players' utilities depend on the actions of the two players. The action of the first player may reveal to the second player some of his private information. We use the framework of signalling game to solve the first type of problem and that of Bayesian game to solve the second. Other problems include pricing access to the Internet with partial information [52] (by I. Brunetti (Univ. Bologna, Italy), M. Haddad (Univ. Avignon/LIA) and E. Altman). In [45], M. Haddad and E. Altman, in collaboration with P. Wiecek (Wroclaw Univ. of Technology, Poland), apply Bayesian games for the association problem in which users have to decide to which access point to connect.

5.3.7. Jamming

We have been working on various models that capture different aspects of jamming (on purpose noise generation). Jamming with partial information is studied in [51] using Bayesian games, by M. Haddad (Univ. Avignon/LIA), E. Altman and S. Azad, as well as [62] and [63] by E. Altman in collaboration with A. Garnaev (St. Petersburg State Univ., Russia) and Y. Hayel (Univ. Avignon/LIA). With K. Avrachenkov, they further consider a dynamic jamming problem in [61]. In all these models the jammer creates interference to the data packets. In [29] V. Kavitha and R. El-Azouzi (Univ. Avignon/LIA), R. Sundaresan (IISc, Bangalore, India), and E. Altman study a different type of jamming game. The jammer attacks the signalling channel and not the data itself. A Bayesian game is formulated and solved there.

5.3.8. Network neutrality and collusions

Network neutrality is a key issue in the future Internet. It is related to the question of whether the access to Internet will remain a universal service or whether it would be regulated by market forces according to economic interests of those that control the Internet access. One form of network non-neutrality is when an ISP gives preferential treatment to one content provider over others. We call this "collusion" or "vertical monopoly". In collaboration with T. Jimenez and Y. Hayel (Univ. Avignon/LIA), E. Altman studies this in [71] along with "horizontal monopolies" that may occur when several ISPs merge. They introduce a new concept of "price of collusion" and identify in [44] cases in which not only consumers loose from collusions but also the colliding agents, as also seen in a different model for network non-neutrality given in [69] by M. K. Hanawal (also with Univ. Avignon/LIA) and E. Altman in collaboration with R. Sundaresan (IISc, Bangalore, India). This is related to a special kind of Braess type paradox.

5.3.9. Competition over popularity in social networks

We focus on competition of video contents for popularity. We analyze the impact of sharing, embedding, advertisement and other actions by the users for increasing the popularity and visibility. This then allowed E. Altman in [80], [38] and [95] to propose stochastic game models and to fully determine the equilibrium policy. He further proposes a dynamic game for the study of partial information and obtain the equilibrium policies and equilibrium performance. In [39], [79] the results are further extended for the wireless context.

5.3.10. Stochastic geometry methods for wireless design issues

Stochastic geometry seems to be the adequate tool in order to model correctly randomness in the location of networks elements such as the mobile terminals and the fixed base stations. Modeling the locations of both as independent spatial processes, In [66] and [25], M. K. Hanawal and E. Altman study in collaboration with F. Baccelli (TREC, INRIA - ENS) properties of Nash equilibria obtained in a multiple access game. They also derive the saddle point obtained in jamming games [68].

5.3.11. In which content to specialize

E. Altman considers in [40] the question of how should a content provider decide in which content to specialize. He shows that the problem is equivalent to the so called "Crowding" games, which allows him to prove the existence of a pure equilibrium. The conclusion is then that there is no gain by diversifying in several contents.

5.3.12. Cognitive radio

In collaboration with J. Elias (Univ. Paris Descartes-Sorbonne) and F. Martignon (LRI-Univ. Paris-Sud), E. Altman study in [56] the question of which priority level to use in a cognitive radio network: higher priority (primary user) or lower one (secondary user). The utilities are function of both the price and the quality of service. After deriving an equilibrium in this game problem, the authors study the question of how to choose prices so as to induce efficient equilibria.

5.3.13. Constrained games

In collaboration with A. Galindo-Serrano and L. Guipponi (CTTC, Spain), E. Altman studies in [60] a game theoretical problem of power control in several base stations with a coupled constraint: the interference at a given point in space should be upper bounded by some constant. The authors establish the existence of a continuum of constrained equilibria to this type of games and show that there is a unique one with some desirable scaling properties (i.e. that consitutes a normalized Nash equilibrium).

5.3.14. Dynamic coalition games

In collaboration with M. K. Panda and T. Chahed (Telecom SudParis, France), E. Altman considers the question of whether to join a multicast session or not. In contrast to many queueing problems, the congestion here is a desirable property, since the cost per user decreases as the number of users connected to the multicast session increases. In [74] the equilibrium policies are derived; these exhibit a surprising structure.

5.3.15. Evolutionary games

The relatively young theory of Evolutionary games considers a large number of interactions between pairs of randomly selected players. It is thus based on a relatively narrow scope in which the one that interacts is the player. In collaboration with Y. Hayel (Univ. Avignon/LIA) and E. V. Belmega (ETIS/ENSEA - Univ. Cergy-Pontoise - CNRS), E. Altman has been developing in [26] an alternative theory of evolutionary game in which a player consists of a group of interacting agents. This is in line with today's understanding of evolution of species (e.g. Dawkins' book "The Selfish Gene" in which the player is the gene of the species). We plan to apply this to energy dependent power control in wireless systems. We also plan to apply these in other areas such as the evolution of languages over social networks, in which some preliminary results (over Twitter) were already obtained in [81] by E. Altman and Y. Portilla (Univ. Avignon/LIA).

5.4. Green networking

Participants: Sara Alouf, Nicaise Choungmo Fofack, Delia Ciullo, Alain Jean-Marie.

5.4.1. Analysis of power saving in cellular networks with continuous connectivity

We have pursued our effort in the analysis of the continuous connectivity mode used in 4G cellular networks. Assuming Poisson traffic at each user, S. Alouf and V. Mancuso (Institute IMDEA Networks, Madrid, Spain) analyze the impact of 3GPP-defined power saving mechanisms on the performance of users with continuous connectivity. Each downlink mobile user's traffic is seen as $M/G/1$ queue, and the base station's downlink traffic as an $M/G/1\,PS$ queue with multiple classes and inhomogeneous vacations. The model is validated through packet-level simulations in [35]; its results show that consistent power saving can be achieved in the wireless access network.

The case of web traffic is investigated in [13] where the same authors, with the participation of N. Choungmo Fofack, perform in addition a sensitivity analysis to assess the impact of model parameters on the performance and cost metrics. It is found that significant power save can be achieved while users are guaranteed to experience high performance. Important outcomes of this work include the need to limit the number of active users in a cell (to less than 350 users – reasonable for 3GPP LTE, 802.16 and HSPA networks) in order to limit the web page download time, and the need to limit the web page size as large pages can dramatically decrease the energy saving. A *green attitude* would be to design web sites with short pages having few embedded objects, enabling cellular operators to use reasonable power save parameters and yet achieve a dramatic cost economy at both base station and mobile user sides, without any quality degradation.

5.4.2. Analysis of base station sleep modes in cellular networks

D. Ciullo, L. Chiaraviglio (INRIA project-team MASCOTTE), M. Ajmone Marsan (Politecnico di Torino, Italy and Institute IMDEA Networks, Spain), M. Mellia and M. Meo (Politecnico di Torino, Italy) study in [78] base station sleep modes. Putting into sleep mode some base stations in periods of low traffic improves the energy efficiency of cellular access networks. Two schemes are considered whether the sleep mode is activated once per day or multiple times per day having progressively fewer active base stations. For both schemes, the optimal base station sleep times are identified according to the traffic. Considering real traffic traces, the study reveals that significant energy saving can be achieved, the actual value strongly depending on the traffic pattern. An important result is that most of the potential savings can be attained with a single daily sleep mode, avoiding the increased complexity coming from the use of multiple sleep modes per day.

5.4.3. Analysis of sleep modes in backbone networks

The case of backbone networks is considered in [86] where L. Chiaraviglio (INRIA project-team MASCOTTE)), D. Ciullo, M. Mellia and M. Meo (Politecnico di Torino, Italy) formulate a theoretical model based on random graph theory. This model allows to estimate the potential gains achievable by adopting sleep modes in fixed networks where some devices consume energy proportionally to the handled traffic. Putting a given fraction of devices in sleep mode reduces the energy these consume but increases the energy consumed by the devices still active due to the additional load these have to handle. The model of [86] allows to predict how much energy can be saved in different scenarios. The results show that sleep modes can be successfully combined with load proportional solutions. However, if the static power consumption component is one order of magnitude less than the load proportional component, then sleep modes are no longer convenient. Thanks to random graph theory, this model gauges the impact of different properties of the network topology.

5.5. Content-oriented systems

Participants: Konstantin Avrachenkov, Nicaise Choungmo Fofack, Delia Ciullo, Philippe Nain, Giovanni Neglia, Marina Sokol.

5.5.1. Performance analysis of peer-assisted Video-on-Demand (VoD) systems

In [88] and [97], D. Ciullo, V. Martina and E. Leonardi (Politecnico di Torino, Italy), M. Garetto (Università di Torino, Italy), and G. L. Torrisi (CNR, Italy) consider peer-assisted Video-on-Demand systems. Some of the essential aspects of such systems are peer churn, bandwidth heterogeneity, and Zipf-like video popularity. The authors propose an analytical framework to tightly characterize the scaling laws for the additional bandwidth that servers must supply to guarantee perfect service, taking into account these essential aspects.

The results in [88] and [97] reveal that the catalog size and the content popularity distribution have a huge effect on the system performance. Also, users' cooperation can effectively reduce the servers' burden for a wide range of system parameters, confirming it as an attractive solution to limit the costs incurred by content providers as the system scales to large populations of users. Moreover, in [89] the same authors provide important hints for the design of efficient peer-assisted VoD systems under server capacity constraints.

5.5.2. Analysis of TTL-based cache networks

N. Choungmo Fofack, P. Nain and G. Neglia, together with D. Towsley (Univ. of Massachusetts at Amherst, USA) introduced in [87] a novel Time-To-Live (TTL) replacement policy to manage a set of documents buffering routers in information-centric networks. The TTL policy assigns a timer to each content stored in the cache and redraws the timer at each content request. In [53] they have showed that this TTL policy is more general than other policies like least frequently used (LRU), first-in-first-out (FIFO) or random (RND) as it mimics their behavior under an appropriate choice of its parameters. While exact formulas for the performance metrics of interest (hit/miss processes) are derived for a linear network and a tree network with one root cache and N leaf caches, for more general networks, an approximate solution is found with relative errors smaller than 10^{-3} and 10^{-2} for exponentially distributed and constant TTLs respectively. It is demonstrated in [53] that the TTL model can be implemented and used to optimize a multi-content cache network under realistic constraints such as the cache size limitation.

5.5.3. CCN interest routing as multi-armed bandit problem

In [49] K. Avrachenkov and P. Jacko (BCAM, Spain) consider Content Centric Network (CCN) interest forwarding problem as a Multi-Armed Bandit (MAB) problem with delays. The authors investigate the transient behaviour of the ϵ-greedy, tuned ϵ-greedy and Upper Confidence Bound (UCB) interest forwarding policies. Surprisingly, for all the three policies very short initial exploratory phase is needed. It is demonstrated that the tuned ϵ-greedy algorithm is nearly as good as the UCB algorithm, commonly reported as the best currently available algorithm. The uniform logarithmic bound for the tuned ϵ-greedy algorithm in the presence of delays is proved. In addition to its immediate application to CCN interest forwarding, the new theoretical results for MAB problem with delays represent significant theoretical advances in machine learning discipline.

In [46] K. Avrachenkov together with L. Cottatellucci and L. Maggi (both from Eurecom, France) consider the choice of CCN Access Points (APs) when CCN APs are wireless base stations. It is assumed that the slow fading channel attenuations follow an autoregressive model. In the single user case, the authors formulate this selection problem as a restless multi-armed bandit problem and propose two strategies to dynamically select a band at each time slot. The objective is to maximize the SNR in the long run. Each of these strategies is close to the optimal strategy in different regimes. In the general case with several users, the authors formulate the problem as a stochastic game with uncountable state space, where the objective is the SINR. Then the authors propose two strategies to approximate the best response policy for one user when the other users' strategy is fixed.

5.6. Advances in methodological tools

Participants: Eitan Altman, Konstantin Avrachenkov, Alain Jean-Marie, Philippe Nain.

5.6.1. Perturbation analysis

In [17] K. Avrachenkov, together with R. Burachik, J. Filar V. Gaitsgory (Univ. of South Australia, Australia), study a linear programming problem with a linear perturbation introduced through a parameter $\epsilon > 0$. The authors identify and analyze an unusual asymptotic phenomenon in such a linear program. Namely, discontinuous limiting behavior of the optimal objective function value of such a linear program may occur even when the rank of the coefficient matrix of the constraints is unchanged by the perturbation. The authors show that, under mild conditions, this phenomenon is a result of the classical Slater constraint qualification being violated at the limit and propose an iterative, constraint augmentation approach for resolving this problem.

5.6.2. Zero-sum games

In [18] K. Avrachenkov, together with L. Cottatellucci and L. Maggi (both from Eurecom, France), study zero-sum two-player stochastic games with perfect information. The authors propose two algorithms to find the uniform optimal strategies and one method to compute the optimality range of discount factors. The convergence in finite time for one algorithm is proved. In particular, the uniform optimal strategies are also optimal for the long run average criterion and, in transient games, for the undiscounted criterion as well.

5.6.3. Approximations in semi-Markov zero-sum games

In conjunction with E. Della Vecchia and S. Di Marco (both from National Univ. Rosario, Argentina), A. Jean-Marie has pursued the studies on the Rolling Horizon procedure and other approximations in stochastic control problems. Their first study on convergence conditions for average-cost MDPs has been published in [23].

They have then turned to the case of discounted semi-Markov zero-sum games. Generalizing previous contributions of the literature, they have established existence conditions and geometric convergence results when action spaces are compact and rewards possibly unbounded. The bounds they obtain hold for the Rolling Horizon procedure as well as for variants called Approximate Rolling Horizon [91]. In the same semi-Markovian context, they have also performed a sensitivity analysis of the model with respect to its parameters: cost function, discount factor, transition probabilities and state space [90].

5.6.4. Retrial queues

In [84] K. Avrachenkov and P. Nain, in collaboration with U. Yechiali (Tel Aviv Univ.), consider a retrial system with two input streams and two orbit queues. More specifically, there are two independent Poisson streams of jobs feeding a single-server service system having a limited common buffer that can hold at most one job. If a type-i job (i=1,2) finds the server busy, it is blocked and routed to a separate type-i retrial (orbit) queue that attempts to re-dispatch its jobs at its specific Poisson rate. This creates a system with three dependent queues. Such a queueing system serves as a model for two competing job streams in a carrier sensing multiple access system. The authors study the queueing system using multi-dimensional probability generating functions, and derive its necessary and sufficient stability conditions while solving a boundary value problem. Various performance measures are calculated and numerical results are presented.

5.6.5. Branching processes

In collaboration with D. Fiems (Gent Univ., Belgium), E. Altman introduces in [41] non-standard new branching processes and applies them to evaluate queueing processes. The processes are characterized by replacing the standard Algebra involved in the definition of branching processes by the max-plus algebra. Among the applications introduced are (i) polling systems with infinite server, and (2) new Cruz type bounds for systems with feedback.

Standard branching have been used in the past to study polling systems. In [30] V. Kavitha (LIA/Univ. Avignon) and E. Altman have revisited this method and applied it to spatial sensors, that receive or send data via a mobile relay or base stations. They derive conservation laws for this continuous state space polling system which allows them to compute optimal polling strategies.

D. Fiems (Gent Univ., Belgium) and E. Altman have further used in [24] semi-linear processes, which extend branching processes, to compute expected waiting times in polling systems with generally distributed walking times (the standard i.i.d. assumption is replaced with the assumption that the walking times are stationary ergodic).

In [22], the problem of parallel TCP connections is studied by O. Czerniak and U. Yechiali (Tel Aviv Univ., Israel), in collaboration with E. Altman, for a model in which, when the sum of throughputs reaches some value, there is a loss. It is assumed that the connection to suffer the loss is chosen according to a round robin policy. The expected throughputs of the connections are computed using an approach based on multitype branching processes.

MASCOTTE Project-Team

6. New Results

6.1. Network Design and Management

Participants: Gianlorenzo D'Angelo, Jean-Claude Bermond, Khoa Phan, David Coudert, Frédéric Giroire, Joanna Moulierac, Nicolas Nisse, Ronan Pardo Soares, Stéphane Pérennes, Issam Tahiri.

6.1.1. Network Design

Network design is a very wide subject that concerns all kinds of networks. We mainly study telecommunications networks which can be either physical networks (backbone, access, wireless, ...) or virtual (logical) ones. The objective is to design a network able to route a (given, estimated, dynamic, ...) traffic under some constraints (e.g. capacity) and with some quality of service (QoS) requirements. Usually the traffic is expressed as a family of requests with parameters attached to them. In order to satisfy these requests, we need to find one (or many) path(s) between their end nodes. The set of paths is chosen according to the technology, the protocol or the QoS constraints. The design can be done at the conception of the network (i.e. when conceiving a virtual network in MPLS where we have to establish virtual paths) or to adapt the network to changes (failures, new link, updates of routers, variation of traffic, ...). Finally there are various optimization criteria which differ according to the point of view: for a network user they are related to his/her satisfaction (minimizing delays, increasing available bandwidth, ...), while for a network operator, economics criteria like minimizing deployment and operating costs are more important.

This very wide topic is addressed by a lot of academic and industrial teams in the world. Our approach is to attack these problems with tools from Discrete Mathematics.

6.1.1.1. All-Optical Label Switching, AOLS

All-Optical Label Switching (AOLS) is a promising technology that performs packet forwarding without any optical-electrical-optical conversions, thus speeding up the forwarding. However, the cost of this technology requires limiting the number of labels needed to ensure the forwarding when routing a set of requests using GMPLS technology. In particular, this prevents the usage of label swapping techniques.

We have studied the routing problem in this context using label stacking techniques. We have formalized the problem by associating to each routing strategy a logical hypergraph, called a hypergraph layout, whose hyperarcs are dipaths of the physical graph, called tunnels in GMPLS terminology. We defined a cost function for the hypergraph layout, depending on its total length plus its total hop count. Minimizing the cost of the design of an AOLS network can then be expressed as finding a minimum cost hypergraph layout. In [24], we prove hardness results for the problem. On the other hand, we provide approximation algorithms, in particular an $O(\log n)$-approximation for symmetric directed networks. We focused on the case where the physical network is a directed path, providing a polynomial-time dynamic programming algorithm first for one source, and then for a fixed number k of sources running in time $O(n^{k+2})$.

6.1.1.2. Protocols

IP multicast is a protocol that deals with group communications with the aim of reducing traffic redundancy in the network. However, due to difficulty in deployment and poor scalability with a large number of multicast groups, IP multicast is still not widely deployed nor used on the Internet. Recently, Xcast6 and Xcast6 Treemap, two network layer multicast protocols, have been proposed with complementary scaling properties to IP multicast: they support a very large number of active multicast sessions. However, the key limitation of these protocols is that they only support small multicast groups. To overcome this limitation, we have proposed the Xcast6 Treemap Island [59], [60], a hybrid model of Application Layer Multicast (ALM) and Xcast6 that can work for large multicast groups. We have shown the feasibility of our model by simulation and comparison with IP multicast and NICE protocols.

Congestion control is a distributed algorithm to share network bandwidth among competing users on the Internet. In the common case, quick response time for mice traffic (http traffic) is desired when mixed with elephant traffic (ftp traffic). The current approach using loss-based with Additive Increase, Multiplicative Decrease (AIMD) is too greedy and eventually, most of the network bandwidth would be consumed by elephant traffic. As a result, it causes longer response time for mice traffic because there is no room left at the routers. MaxNet is a new TCP congestion control architecture using an explicit signal to control transmission rate at the source node. In [60] , we show that MaxNet can control well the queue length at routers and therefore the response time to http traffic is several times faster than with TCP Reno/RED.

6.1.1.3. Shared Risk Link Group

The notion of *Shared Risk Link Group*, SRLG has been introduced to capture multiple correlated failures in a network. A SRLG is a set of links that fail simultaneously if a given event (risk) occurs. In such multiple failures scenario, the problem of Diverse Routing consists in finding two SRLG-disjoint paths between a pair of nodes. We consider in [42], [66] such problem for localized failures, when all the links of a SRLG verify the star property i.e. when they are incident to the same node. We prove that in this case the problem is in general NP-complete and determine some polynomial cases.

6.1.1.4. Data Gathering in Radio Networks

We study the problem of gathering information from the nodes of a radio network into a central node. We model the network of possible transmissions by a graph and consider a binary model of interference in which two transmissions interfere if the distance in the graph from the sender of one transmission to the receiver of the other is d_I or less. A *round* is a set of non-interfering transmissions. In [25], we determine the exact number of rounds required to gather one piece of information from each node of a square two-dimensional grid into the central node. The even case uses a method based on linear programming duality to prove the lower bound, and sophisticated algorithms using the symmetry of the grid and non-shortest paths to establish the matching upper bound. We then generalize our results to hexagonal grids.

Other results on multi-interface networks were obtained outside of MASCOTTE [30], [31], [55].

6.1.2. Routing

The problem of finding and updating shortest paths in distributed networks is considered crucial in today's practical applications. In the recent past, there has been a renewed interest in designing new efficient distance-vector algorithms (e.g., the distributed Bellman-Ford method implemented in the routing information protocol, RIP) as an alternative to link-state solutions (e.g., open shortest path first, OSPF) for large-scale distributed networks such as the autonomous systems topology of the Internet.

This year, we have proposed a new loop-free distance-vector routing algorithm, called LFR (Loop Free Routing), which is able to update the shortest paths of a distributed network with n nodes in fully dynamic scenarios [47]. We compared experimentally this new algorithm with DUAL, one of the most popular loop-free distance vector algorithms which is part of CISCO's EIGRP protocol. Our experiments on CAIDA IPv4 routed /24 topology dataset show that LFR out-performs DUAL in terms of memory requirements and number of messages.

We then proposed a new technique, called Distributed Computation Pruning (DCP) [48], for reducing the total number of messages sent and the space occupancy per node of every distance-vector routing algorithm based on shortest paths. We have evaluated experimentally the combination of DCP with DUAL and with LFR. We have observed that these combinations lead to a significant gain both in terms of number of messages sent and memory requirements per node.

We have also considered routing problems arising in road networs. In particular, we have conducted a theoretical study of the graph-augmentation problem of adding shortcuts in order to speedup route planning techniques [23]. We studied the algorithmic complexity of the problem and proposed approximation algorithms for a special graph class. We have also investigated ILP-based exact approaches and show how to stochastically evaluate a given shortcut assignment on graphs that are too large to do so exactly.

6.1.2.1. Compact routing

With the constant increase of the number of routing entries in the Internet, the size of the routing tables stored at router nodes increases drastically. Routing schemes such as BGP are showing their limits in terms of update time, search time, cost of signaling, etc. and alternatives have to be proposed. In particular, compact routing schemes propose interesting trade-offs between the size of the routing tables and the quality of the routes. They also take advantage of the particular properties arising in large scale networks such as low (logarithmic) diameter and high clustering coefficient.

High clustering coefficient implies the existence of few large induced cycles. Considering this fact, we proposed in [37] a routing scheme that computes short routes in the class of k-chordal graphs, i.e., graphs with no induced cycles of length more than k. Our routing scheme achieves an additive stretch of at most $k-1$, and the routing tables are computed with a distributed algorithm which uses messages of size $O(\log n)$ and takes $O(D)$ time, where D is the diameter of the network.

We also used *cops-and-robber* games (See Section 6.2.1.2) to propose the first compact routing scheme for k-chordal graphs using routing tables, addresses and headers of size $O(\log n)$ bits and achieving an additive stretch of $O(k \log \Delta)$ [58], [57], [77]. This scheme is based on a new structural decomposition for a graph class including k-chordal graphs: we proposed a quadratic algorithm that, given a graph G and $k \geq 3$, either returns an induced cycle larger than k in G, or computes a *tree-decomposition* of G, each *bag* of which contains a dominating path with at most $k-1$ vertices. We thus proved that any k-chordal graph with maximum degree Δ has treewidth at most $(k-1)(\Delta-1)+2$, improving the $O(\Delta(\Delta-1)^{k-3})$ bound of Bodlaender and Thilikos (1997). Moreover, any graph admitting such a tree-decomposition has small hyperbolicity.

In addition, we have pursued our investigation of the kind of structural graph properties that can or cannot be deduced from local (partial) views of the network. Such knowledge is crucial for the design of routing schemes. To this end, we have exhibited a hierarchy of problems and distributed models of computation [40].

6.1.2.2. Routing models evaluation

The evaluation of new routing models asks for large-scale and intensive simulations. However, existing routing models simulators such as DRMSim are limited in terms of the number of routing table entries it can dynamically process and control on a single computer. Therefore, we have conducted a feasibility study of the extension of DRMSim so as to support the Distributed Parallel Discrete Event paradigm [46]. We have studied several distribution models and their associated communication overhead. We have in particular evaluated the expected additional time (in hours) required by a distributed simulation of BGP (border gate protocol), the current interdomain routing protocol of the Internet, compared to its sequential simulation. We show that such a distributed simulation of BGP is possible with a reasonable time overhead.

6.1.2.3. Reconfiguration

In production networks, traffic evolution, failures and maintenance operations force to adapt regularly the current configuration of the network (virtual topology, routing of connections). The routing reconfiguration problem in WDM networks consists of scheduling the migration of established lightpaths from current routing to a new pre-computed one while minimizing service disruptions. We have shown in the past the relations between this problem and the graph searching problem and established NP-completeness and inapproximability results.

This year, we proved the monotonicity of the *process strategy* game [78], the graph searching game modeling the routing reconfiguration problem. Then, we have investigated on the influence of physical layer impairment constraints on the reconfiguration problem [41]. Setting up a new wavelength in a fiber of a WDM network requires recalibrating the other wavelengths passing through this fiber. This induces a cost (e.g., time, energy, degradation of QoS) that depends nonlinearly on the number of wavelengths using the fiber. Therefore, the order in which requests are switched affects the total cost of the operation. We have studied the corresponding optimization problem by modeling the cost of switching a request as a non-linear function depending on the load of the links used by the new lightpath. We have proved that determining the optimal rerouting order is NP-complete for a 2-nodes network, established general lower and upper bounds, identified classes of instances where the problem can be solved in polynomial time, and proposed a heuristic algorithm.

6.1.3. Energy efficiency

Recently, energy-aware routing has gained increasing popularity in the networking research community. The idea is that traffic demands are aggregated over a subset of the network links, allowing other links to be turned off to save energy. We develop several methods to improve routing protocols for backbone, wireless and content delivery networks. Several studies exhibit that the traffic load of the routers only has a small influence on their energy consumption. Hence, the power consumption in networks is strongly related to the number of active network elements, such as interfaces, line cards, base chassis,... The goal thus is to find a routing that minimizes the (weighted) number of active network elements used when routing. In [62], we exhibit that the power consumption can be reduced of approximately 33 MWh for a medium-sized backbone network.

In [54], we propose GreenRE - a new energy-aware routing model with the support of the new technique of data redundancy elimination (RE). Based on real experiments on Orange Labs platform and on simulations on several network topologies, we show that GreenRE can gain further 30% energy savings in comparison with the traditional energy-aware routing model.

One of the new challenges facing research in wireless networks is the design of algorithms and protocols that are energy aware. In [33], we use for the first time the evolving graph combinatorial model as a tool to prove an NP-Completeness result, namely that computing a Minimum Spanning Tree of a planar network in the presence of mobility is actually NP-Complete.

Recently, there is a trend to introduce content caches as an inherent capacity of network equipment, with the objective of improving the efficiency of content distribution and reducing network congestion. In [63], we study the impact of using in-network caches and CDN cooperation on an energy-efficient routing: up to 23% of power can be saved in the backbone this way.

In [32], we study the energy efficiency of the networking part of data centers, accounting for between 10-20% of the total power consumption. We proposed a novel approach, called VMPlanner, for power reduction in the virtualization-based data centers. The idea of VMPlanner is to optimize both virtual machine placement and traffic flow routing so as to turn off as many unneeded network elements as possible for power saving.

Finally, in [56], [38], we summarize the main research results of the last years for energy efficiency for backbone, wireless, cellular and content distribution networks and highlight the main challenges of the field. Results are given for two operator networks, considering power and traffic forecasts for 2020.

6.2. Graph Theory

Participants: Julio Araújo, Jean-Claude Bermond, Frédéric Giroire, Frédéric Havet, František Kardoš, Ana Karolinna Maia, Remigiusz Modrzejewski, Leonardo Sampaio, Michel Syska.

6.2.1. Algorithms in graphs

MASCOTTE is also interested in the algorithmic aspects of Graph Theory. In general we try to find the most efficient algorithms to solve various problems of Graph Theory and telecommunication networks.

6.2.1.1. Complexity and Computation of Graph Parameters

We used graph theory to model various networks' problems. In general we study their complexity and then we investigate the structural properties of graphs that make these problems hard or easy. In particular, we try to find the most efficient algorithms to solve the problems, sometimes focusing on specific graph classes where the problems are polynomial-time solvable.

Degree Constraint Subgraphs. A natural question in current social networks is *How do one find a small community (subgraph) in which anyone as at least* d *friends (neighbors)?* This problem can be modelled as degree-constrained subgraph problems where the objective is to find an optimal weighted subgraph, subject to certain degree constraints (in which each node has degree at most d), in a weighted graph. When $d = 2$, the problem is easy to solve since one simply needs to compute the girth of the graph. In [16], we proved that the problem is not in Apx for any $d \geq 3$. The proof is obtained by a reduction from Vertex Cover in regular graphs, followed by the use of an error amplification technique. On the positive side, we give an $\frac{n}{\log n}$-approximation algorithm for the class of graphs excluding a fixed graph H as a minor (including planar or bounded genus graphs), using dynamic programming.

Hyperbolicity in Large graphs. Hyperbolicity is a geometric notion that measure how the various shortest paths connecting two vertices can diverge in a graph. Knowing its value provides information on the geometry of the network, moreover it has practical implications for shortest path routing. Hyperbolicity can be computed in polynomial time algorithm ($\Theta(n^4)$). This is far from being practical for large graphs. So, in [69] we proposed a scalable algorithm for this problem. We also led some computational experiments of our algorithms on large-scale graphs.

Hull Number of graphs. In [64], we study the (geodesic) hull number of graphs. For any two vertices $u, v \in V$ of a connected undirected graph $G = (V, E)$, the closed interval $I[u, v]$ of u and v is the set of vertices that belong to some shortest (u, v)-path. For any $S \subseteq V$, let $I[S] = \bigcup_{u,v \in S} I[u, v]$. A subset $S \subseteq V$ is (geodesically) convex if $I[S] = S$. Given a subset $S \subseteq V$, the convex hull $I_h[S]$ of S is the smallest convex set that contains S. We say that S is a hull set of G if $I_h[S] = V$. The size of a minimum hull set of G is the hull number of G, denoted by $hn(G)$. First, we show a polynomial-time algorithm to compute the hull number of any P_5-free triangle-free graph. Then, we present four reduction rules based on vertices with the same neighborhood. We use these reduction rules to propose a fixed parameter tractable algorithm to compute the hull number of any graph G, where the parameter can be the size of a vertex cover of G or, more generally, its neighborhood diversity, and we also use these reductions to characterize the hull number of the lexicographic product of any two graphs. More on the hull number of graphs may be found in Araujo's thesis [13].

6.2.1.2. Graph Searching, Cops and Robber Games

Pursuit-evasion encompasses a wide variety of combinatorial problems related to the capture of a fugitive residing in a network by a team of searchers. The goal consists in minimizing the number of searchers required to capture the fugitive in a network and in computing the corresponding capture strategy. This can also be viewed as cleaning the edges of a contaminated graph. We investigated several variants of these games.

Web Caching & the surfer Game. A surprising application of some variant of pursuit-evasion games (namely Cops and Robber games) is the problem for a web-browser to download documents in advance while an internaut is surfing on the Web. In [53], [52], we provide a modelling of the prefetching problem in terms of Cops and Robber games. The parameter to be optimized is then the download-speed necessary for the Internaut only accesses to already download webpages. This allows us to provide several complexity results and polynomial-time algorithms in some graph classes.

Connected Graph Searching. Another variant of pursuit-evasion games is graph searching which is mainly related to graph decompositions. For instance, the minimum number of searchers needed to capture an invisble fugitive in a graph is equal to its pathwidth plus one. In [21], we investigated the connected variant of this game. A strategy is called connected if the clear part (the part where the fugitive cannot stand) always induces a connected subgraph. The main motivation for studying connected graph searching is the design of distributed protocols allowing searchers to compute a capture strategy (see also Section 6.2.1.3). [21] gathers most of the results of the last decade concerning connected graph searching, mainly focussing on the cost of connectivity in terms of number of searchers.

6.2.1.3. Distributed Algorithms

We investigated algorithmic problems arising in complex networks like the Internet or social networks. In this kind of networks, problems are becoming harder or impracticable because of the size and the dynamicity of these networks. One way to handle the dynamicity is to provide (distributed) fault tolerant algorithms. Studying the mobile agents paradigm seems to be a promissing approach (somehow related to Cops and Robber in Section 6.2.1.2) to adress some models of distributed computing. We considered distributed or even self-stabilizing algoritms for gathering and graph searching problems.

Graph Searching and Routing Reconfiguration. In [29], we developed a generic distributed algorithm for computing and updating various parameters on trees including the process number (see Section 6.1.2.3), and other related graph searching parameters (see Section 6.2.1.2). We also proposed an incremental version of the algorithm allowing to update these parameters after addition or deletion of any tree edge.

Robots in anonymous networks. Motivated by the understanding of the limits of distributed computing, we consider a recent model of robot-based computing which makes use of identical, memoryless mobile robots placed on nodes of anonymous graphs. The robots operate in Look-Compute-Move cycles that are performed asynchronously for each robot. In particular, we consider various problems such as graph exploration, graph searching and gathering in various graph classes. We provide a new distributed approach which turns out to be very interesting as it neither completely falls into symmetry-breaking nor into symmetry-preserving techniques. More precisely, we design algorithms for the gathering in rings [51], [70], grid [50] and trees [61]. We also proposed a general approach [71] to solve the three problems in rings. Finally, in [67], [44], [43], algorithms are designed to solve the graph searching problem in trees.

6.2.2. Structural graph theory

6.2.2.1. Directed graphs

Graph theory can be roughly partitioned into two branches: the areas of undirected graphs and directed graphs (digraphs). Even though both areas have numerous important applications, for various reasons, undirected graphs have been studied much more extensively than directed graphs. One of the reasons is that many problems for digraphs are much more difficult than their analogues for directed graphs. For example, one of the cornerstones of modern (undirected) graph theory is Minor Theory of Robertson and Seymour. Unfortunately, we cannot expect an equivalent for directed graphs. Minor Theory implies in particular that, for any fixed H, detecting a subdivision of H in an input graph G can be performed in polynomial time by the Robertson and Seymour linkage algorithm. In contrast, the analogous subdivision problem for digraph can be either polynomial-time solvable or NP-complete, depending on the fixed digraph H. In [65], we give a number of examples of polynomial instances, several NP-completeness proofs as well as a number of conjectures and open problems. We also investigated the related problem in which we want to detect an *induced* subdivision of H. Already, for undirected graphs the complexity of this problem depends on H. In [20], we show that for digraphs the complexity of this problem depends on H and on whether the input digraph G must be an oriented graph or is allowed to contain 2-cycles. We give a number of examples of polynomial instances as well as several NP-completeness proofs.

In a directed graph, a *star* is an arborescence with at least one arc, in which the root dominates all the other vertices. A *galaxy* is a vertex-disjoint union of stars. In [34], we consider the Spanning Galaxy problem of deciding whether a digraph D has a spanning galaxy or not. We show that although this problem is NP-complete (even when restricted to acyclic digraphs), it becomes polynomial-time solvable when restricted to strong digraphs. In fact, we prove that restricted to this class, the Spanning Galaxy problem is equivalent to the problem of deciding if a strong digraph has a strong digraph with an even number of vertices. We then show a polynomial-time algorithm to solve this problem. We also consider some parameterized versions of the Spanning Galaxy problem. Finally, we improve some results concerning the notion of *directed star arboricity* of a digraph D, denoted $dst(D)$, which is the minimum number of galaxies needed to cover all the arcs of D. We show in particular that $dst(D) \leq \Delta(D) + 1$ for every digraph D and that $dst(D) \leq \Delta(D)$ for every acyclic digraph D.

Hypergraphs are a generalization of graphs, in which every edge is incident to a set of vertices of any size (not necessarily 2). Like for digraphs, a lot fewer is known about them than about graphs. The two notions of eulerian and hamoltinians cycles have been extensively studied for graphs and digraphs. The analogue notion of eulerian cycle in a hypergraph was only introduced in 2010 by Lonc and Naroski. In [72], we introduce the notions of eulerian and hamiltonian circuits in directed hypergraphs. We show that both associated decision problems are NP-complete. Some necessary conditions for a dihypergraph to be have an eulerian circuit are presented. We exhibit some families of hypergraphs for which those are sufficient conditions. We also generalize a part of the properties of eulerian digraphs to the uniform and regular directed hypergraphs. Finally, we show that the de Bruijn and Kautz dihypergraphs are eulerian and hamiltonian in most cases.

6.2.2.2. Graph colouring

We mainly study graph colouring problems that model channel assignment problems.

A well-known such general problem is the following: we are given a graph G, whose vertices correspond to transmitters, together with an edge-weghting w. The weight of an edge corresponds to the minimum separation between the channels on its endvertices to avoid interferences. (If there is no edge, no separation is required, the transmitters do not interfere.) We need to assign positive integers (corresponding to channels) to the vertices so that for every edge e the channels assigned to its endvertices differ by at least $w(e)$. The goal is to minimize the largest integer used, which corresponds to minimizing the *span* of the used bandwidth.

We mainly studied a particular, yet quite general, case, called *backbone colouring*, in which there are only two levels of interference. So we are given a graph G and a subgraph H, called *the backone*. Two adjacent vertices in H must get integers at least q apart, while adjacent vertices in G must get integers at distance at least 1. The minimum span is this case is called the q-backbone chromatic number and is denoted $BBC_q(G,H)$. Backbone forests in planar graphs are of particular interests. In [74], we give a series of NP-hardness results as well as upper bounds for $BBC_q(G,H)$, depending on the type of the forest (matching, galaxy, spanning tree). Eventually, we discuss a circular version of the problem. In [73], we also consider a list version of the problem in which every vertex must be assigned an integer in its own list of available ones. We provide bounds using the Combinatorial Nullstellensatz for the list version on the channel assignment problem. Through this result and through structural approaches, we obtain good upper bounds for forests and matching backbone in planar graphs. In [68], we give an evidence to a conjecture of Broersma et al. stating that $BBC_2(G,T) \leq 6$, for every planar graph G and spanning tree T. We prove this conjecture in the particular case when T has diameter at most 4.

Another meaningful and very well-studied particular case of backbone colouring is $L(p,1)$-labelling, which is p-backbone colouring of (G^2,G), where G^2 is the square of G (the graph with same vertex set as G, in which two vertices are adjacents if they are at distance at most 2 in G). Griggs and Yeh conjecture in 1992, that for every graph with maximum degree $\Delta \geq 2$, $BBC_2(G^2,G) \leq \Delta^2+1$. In [36], we prove this conjecture when Δ is large. In fact, we prove a more general statement. We prove for any q and sufficiently large Δ, if $\Delta(H) \leq \Delta^2$ and $\Delta(G) \leq \Delta$, then $BBC_q(H,G) \leq \Delta^2+1$. Our result also holds for the list version.

In [17], we studied another colouring problem motivated by a practical frequency assignment problem and, up to our best knowledge, new. In wireless networks, a node interferes with other nodes, the level of interference depending on numerous parameters: distance between the nodes, geographical topography, obstacles,... We model this with a weighted graph (G,w) where the weight function w on the edges of G represents the noise (interference) between the two end-vertices. The total interference in a node is the sum of all the noises of the nodes emitting on the same frequency. A *weighted t-improper k-colouring* of (G,w) is a k-colouring of the nodes of G (assignment of k frequencies) such that the interference at each node does not exceed the threshold t. We consider the Weighted Improper Colouring problem which consists in determining the *weighted t-improper chromatic number* defined as the minimum integer k such that (G,w) admits a weighted t-improper k-colouring. We also consider the dual problem, denoted the Threshold Improper Colouring problem, where, given a number k of colours, we want to determine the minimum real t such that (G,w) admits a weighted t-improper k-colouring. We show that both problems are NP-hard and present general upper bounds for both problems; in particular we show a generalisation of Lovász's Theorem for the weighted t-improper chromatic number. Motivated by the original application, we study a special interference model on various grids (square, triangular, hexagonal) where a node produces a noise of intensity 1 for its neighbours and a noise of intensity 1/2 for the nodes at distance two. We derive the weighted t-improper chromatic number for all values of t.

Since some of the channel assignment problems must be done on-line, we are interested in some on-line graph colouring heuristics. We only studied such heuristics for the classical proper colouring. The easiest one, and the most widespread one, is the greedy algorithm, which colours the vertices one after another, giving to each vertex the smallest possible positive integer that is not already used by one of its neighbours. The *Grundy number* of a graph G is the largest number of colours used by any execution of the greedy algorithm to colour G. In [27], we give new bounds on the Grundy number of the different product of two graphs. The problem of determining the Grundy number of G is polynomial-time solvable if G is a P_4-free graph and NP-hard if G is a P_5-free graph. In [19], we define a new class of graphs, the *fat-extended P_4-laden graphs*, and we show a polynomial-time algorithm to determine the Grundy number of any graph in this class. Our class intersects the class of P_5-free graphs and strictly contains the class of P_4-free graphs. More precisely, our result implies that the Grundy number can be computed in polynomial time for any graph of the following classes: P_4-reducible, extended P_4-reducible, P_4-sparse, extended P_4-sparse, P_4-extendible, P_4-lite, P_4-tidy, P_4-laden and extended P_4-laden, which are all strictly contained in the fat-extended P_4-laden class.

A colouring c of a graph $G=(V,E)$ is a *b-colouring* if in every colour class there is a vertex whose neighborhood intersects every other colour classes. Such a colouring appears, when we try to optimize on-line the colouring of a graph, by changing the colour of all vertices of a colour class if it is possible. The *b-chromatic number* of G, denoted $\chi_b(G)$, is the greatest integer k such that G admits a b-coloring with k colours. A graph G is *tight* if it has exactly $m(G)$ vertices of degree $m(G)-1$, where $m(G)$ is the largest integer m such that G has at least m vertices of degree at least $m-1$. Determining the b-chromatic number of a tight graph had been shown to be NP-hard even for a connected bipartite graph. In [35], we show that it is also NP-hard for a tight chordal graph, and that the b-chromatic number of a split graph can be computed in polynomial time. Then we define the b-closure and the partial b-closure of a tight graph, and use these concepts to give a characterization of tight graphs whose b-chromatic number is equal to $m(G)$. This characterization is used to develop polynomial-time algorithms for deciding whether $\chi_b(G)=m(G)$, for tight graphs that are complement of bipartite graphs, P_4-sparse and block graphs. We generalize the concept of pivoted tree introduced by Irving and Manlove and show its relation with the b-chromatic number of tight graphs.

Many more results on greedy colourings and b-colourings have been proved in Sampaio's thesis [14].

We studied other variations of graph colouring. In [18], we aim at characterizing the class of graphs that admit a good edge-labelling. Such graphs are interesting, as they correspond to set of requests in UPP-digraphs (in which there is at most one dipath from a vertex to another) for which the minimum number of wavelengths is equal to the maximum load. This implies that the problem can be solved efficiently. First, we exhibit infinite families of graphs for which no good edge-labelling can be found. We then show that deciding if a graph admits a good edge-labelling is NP-complete. Finally, we give large classes of graphs admitting a good edge-labelling: C_3-free outerplanar graphs, planar graphs of girth at least 6, subcubic $\{C_3, K_{2,3}\}$-free graphs.

For a connected graph G of order at least 3 and a k-labelling $c: E(G) \to \{1, 2, \cdots, k\}$ of the edges of G, the *code* of a vertex v of G is the ordered k-tuple $(n_1, \cdots, n_k)$, where n_i is the number of edges incident with v that are labelled i. The k-labelling c is detectable if every two adjacent vertices of G have distinct codes. The minimum positive integer k for which G has a detectable k-labelling is the *detection number* of G. In [76], we show that it is NP-complete to decide if the detection number of a cubic graph is 2. We also show that the detection number of every bipartite graph of minimum degree at least 3 is at most 2. Finally, we give some sufficient condition for a cubic graph to have detection number 3.

MESCAL Project-Team

6. New Results

6.1. Analysis and Control of Large Stochastic Systems

Perfect sampling is a very efficient technique that uses coupling arguments to provide a sample from the stationary distribution of a Markov chain in a finite time without ever computing the distribution. Even though, the general (non-monotone) case needs to consider the whole state space, we developed a new approach for the general case that only needs to consider two trajectories, an approach which is particularly effective when the state space can be partitioned into pieces where envelopes can be easily computed [8]. Importantly, we also showed that perfect sampling is possible in Jackson networks, even though the underlying Markov chain has a large or even infinite state space and illustrated the efficiency of our approach via numerical simulations [17]. In a similar vein, given that the analysis of a system's dynamics relies on the collection and the description of events, we developed in [37] a new approach to reduce the descriptional complexity of a system by aggregating events' properties, such as their Shannon entropy, entropy gain, divergence etc. These measures were applied to the evaluation of geographic aggregations in the context of news analysis and they allowed us to determine which abstractions one should prefer depending on the task to perform.

In the study of Markov decision processes composed of a large number of objects, we showed that the optimal reward satisfies a Bellman equation, which converges to the solution of a continuous Hamilton-Jacobi-Bellman (HJB) equation based on the mean field approximation of the Markov decision process [10]. We also gave bounds on the difference of the rewards and an algorithm for deriving an approximating solution to the Markov decision process from a solution of the HJB equations. Furthermore, we also studied deterministic limits of Markov processes with discontinuous drifts and showed that under mild assumptions, the stochastic system is a constant-step stochastic approximation algorithm which converges to a differential inclusion obtained by convexifying the rescaled drift of the Markov chain [9].

Finally, in terms of performance evaluation and its applications, we also studied resource-aware business process models by defining a new framework that allows the generation of analytical models. We showed that the analysis of the generated SAN model provides several performance indices we showed that these indices can be easily calculated by a business specialist with no skills in stochastic modeling [7].

6.2. Game Theory and Applications

As far as results in pure game theory are concerned, we studied in [12] a general framework of systems wherein there exists a Pareto optimal allocation that is Pareto superior to an inefficient Nash equilibrium and defined a 'Nash proportionately fair' Pareto optima. In this context, we provided conditions for the existence of a Pareto-optimal allocation that is, truly or most closely, proportional to a Nash equilibrium – an approach with applications in non-cooperative flow-control problems in communication networks.

In a learning context, we also explored what happens beyond the standard first-order framework of continuous time game dynamics and introduced in [42] a class of higher order game dynamics, extending all first order imitative dynamics, and, in particular, the replicator dynamics to higher orders. In stark contrast to the first order case, we showed that weakly dominated strategies become eliminated in all n-th order payoff-monotonic dynamics for all $n > 1$ and strictly dominated strategies become extinct in n-th order dynamics n orders as fast as in first order. Finally, we also established a higher order analogue of the folk theorem of evolutionary game theory which shows that higher order accelerate the rate of convergence to equilibria in games.

In terms of applications, we also examined the distribution of traffic in networks whose users try to minimise their delays by adhering to a simple learning scheme inspired by the replicator dynamics of evolutionary game theory. A major challenge occurs in this context when the users' delays fluctuate unpredictably due to random external factors, but we showed that if users are not too greedy in their learning scheme, then the long-term averages of the users' traffic flows converge to the vicinity of an equilibrium [43].

6.3. Wireless networks

Power and energy considerations in wireless networks have brought to the forefront the need for efficient power allocation and handover policies.

In [13], we analyze the power allocation problem for orthogonal multiple access channels by means of a non-cooperative potential game in which each user distributes his power over the channels available to him. When the channels are static, we show that this game possesses a unique optimum point; moreover, if the network's users follow a distributed learning scheme based on the replicator dynamics of evolutionary game theory, then they converge to this optimum exponentially fast.

On the other hand, in case the network users have access to multiple-antenna technologies (as most smarphone users do nowadays, we also analyze the problem of finding the optimal signal covariance matrix for MIMO multiple access channels by using an approach based on "exponential learning" – a novel optimization method which applies more generally to (quasi-)convex problems defined over sets of positive-definite matrices (with or without trace constraints) [24]. Furthermore, by using a Riemannian-geometric approach, we devise a distributed optimization algorithm which converges to the optimum signal distribution exponentially fast: users attain an ϵ-neighborhood of the system's optimum configuration in time which is at most $\mathcal{O}(\log(1/\epsilon))$ (and, in practice, within only a few iterations, even for large numbers of users) [25].

In the context of heterogeneous wireless networks where vertical handovers are allowed, we also studied a control problem for a new joint admission and resource allocation controller. To account for multi-objective optimization, we considered the maximization of an objective subject to a set of constraints, and we turned this constrained problem into an unconstrained one that we solved numerically using the Semi-Markovian Decision Process (SMDP) framework [19].

6.4. Scheduling

The parallel computing platforms available today are increasingly larger, so it is necessary to develop efficient strategies providing safe and reliable completion for parallel applications. In [6], we proposed a performance model that expresses formally the checkpoint scheduling problem by exhibiting the tradeoff between the impact of the checkpoints operations and the lost computation due to failures. In particular, we proved that the checkpoint scheduling problem is NP-hard even in the simple case of uniform failure distribution and also presented a dynamic programming scheme for determining the optimal checkpointing times in all variants of the problem. On a similar issue, we proposed in [35] a fair scheduling algorithm that handles the problem of fair scheduling by adopting processor fair-share as a strategy for user fairness. Our approach showed that a parallel machine can give a similar type of performance guarantee as a round-robin scheduler, without requiring job preemption been required.

From a network calculus perspective, we presented in [16] a new formalism for data packetization in networks, the “packet curves”. Indeed, a more precise knowledge of the packet characteristics can be efficiently exploited to get tighter performance bounds, for example for aggregation of flows, packet-based service policies and shared buffers; finally, we also gave a model for a wormhole switch and showed how our results can be used to get efficient delay bounds.

6.5. Multi-Core Systems

Modern multi-core platforms feature complex topologies with different cache levels and hierarchical memory subsystems, so thread and data placement become crucial to achieve good performance. In [14], we evaluate CPU and memory affinity strategies for numerical scientific multithreaded benchmarks on multi-core platforms and analyzed hardware performance event counters in order to acquire a better understanding of such impact. Likewise, thread mapping is an appealing approach to efficiently exploit the potential of modern chip-multiprocessors, so we proposed in [18] a dynamic thread mapping approach to automatically infer a suitable thread mapping strategy for transactional memory applications composed of multiple execution phases with potentially different transactional behavior in each phase. Our results showed that the proposed dynamic approach presents performance improvements up to 31% compared to the best static solution. esp From an optimization perspective, the asymmetry in memory access latencies may reduce the overall performance of the system. Therefore, to achieve scalable performance in this environment, we exploited in [28] the machine architecture while taking into account the application communication patterns. Specifically, we introduced a topology-aware asymptotically optimal load balancing algorithm named HwTopoLB which combines the machine topology characteristics with the communication patterns of the application to equalize the application load on the available cores while reducing latencies. We also introduced in [27] a topology-aware load balancer called NucoLB that focuses on redistributing work while reducing communication costs among and within compute nodes, thus leading to performance improvements of up to 20% when compared to state-of-the-art load balancers.

6.6. Cloud Computing

Even though a new era of Cloud Computing has emerged, the characteristics of Cloud load in data centers is not perfectly clear. In [20], we characterized the job/task load and host load in a real-world production data center at Google Inc. by using a detailed trace of over 25 million tasks across over 12,500 hosts. We found that the Google data center exhibits finer resource allocation with respect to CPU and memory than that of Grid/HPC systems and Google jobs are always submitted with much higher frequency and they are much shorter than Grid jobs, leading to higher variance and noise. Moreover, as far as prediction is concerned, we designed in [21] a Bayes model to predict the mean load over a long-term time interval, as well as the mean load in consecutive future time intervals. Real-world experiments showed that our Bayes method achieved high accuracy with a mean squared error of 0.0014 and that it improves the load prediction accuracy by 5.6-50% compared to other state-of-the-art methods based on moving averages, auto-regression, and/or noise filters.

In a similar vein, the exploitation of Best Effort Distributed Computing Infrastructures (BE-DCIs) allows operators to maximize the utilization of the infrastructures, and users to access the unused resources at relatively low cost. Profiling the execution of Bag-of-Tasks (BoT) applications on several kinds of BE-DCIs demonstrates that their task completion rate drops near the end of the execution. In [33], we presented the SpeQuloS service which enhances the QoS of BoT applications executed on BE-DCIs by reducing the execution time, improving its stability, and reporting to users a predicted completion time. We presented the design and development of the framework and several strategies to decide when and how Cloud resources should be provisioned; moreover, performance evaluation using simulations showed that SpeQuloS fulfill its objectives in speeding up the execution of BoTs, in the best cases by a factor greater than 2, while offloading less than 2.5% of the workload to the Cloud. These topics were also further explored in the book chapter [30].

6.7. Experimentation and Visualization in Large Systems

Despite a widespread belief regarding the simulation of large-scale computing systems, we showed in [15] that achieving high scalability does not necessarily require to resort to overly simple models and ignore important phenomena. In fact, by relying on a modular and hierarchical platform representation while taking advantage of regularity when possible, we were able to model systems such as data and computing centers, peer-to-peer networks, grids, or clouds in a scalable way. Finally, in [34], we examined the ability to conduct consistent, controlled, and repeatable large-scale experiments in areas of computer science where availability, repeatability, and open sharing of electronic products are still difficult to achieve.

We also discussed in [22] the concept of the reconstructability of software environments and we proposed a tool for dealing with this problem. In a similar vein, we developed Expo [41], a tool for conducting experiments on distributed platforms. Our experiments confirmed that Expo is a promising tool to help the user with two primary concerns: how to perform a large scale experiment efficiently and easily, together with its reproducilibity.

The exponential number of processes that are executed in high performance applications and the very detailed behavior that we can record about them lead to a trace size explosion both in space and time dimensions. Thus, if the amount of data is not properly treated for visualization, the analysis may give the wrong idea about the behavior registered in the traces. We dealt with this issue in [38] in two ways: first, by detailing data aggregation techniques that are fully configurable by the user to control the level of details in both space and time dimensions, and second, by presenting two visualization techniques that take advantage of the aggregated data to scale.

Furthermore, given that the performance of parallel and distributed applications is highly dependent on the characteristics of the execution environment, the network topology and characteristics directly impact data locality and movements as well as contention. Unfortunately few visualization available to the analyst are capable of accounting for such phenomena, so we proposed in [39] an interactive topology-based visualization technique based on data aggregation that enables to correlate network characteristics, such as bandwidth and topology, with application performance traces. Such visualization techniques enable us to explore and understand non-trivial behaviors that are impossible to grasp otherwise and the combination of multi-scale aggregation and dynamic graph layout allows us to scale the visualization seamlessly to large distributed systems.

MOAIS Project-Team

6. New Results

6.1. Work Stealing inside GPU

Graphics Processing units (GPU) have become a valuable support for High Performance Computing (HPC) applications. However, despite the many improvements of General Purpose GPUs, the current programming paradigms available, such as NVIDIA?s CUDA, are still low-level and require strong programming effort, especially for irregular applications where dynamic load balancing is a key point to reach high performances. We have introduced a new hybrid programming scheme for general purpose graphics processors using two levels of parallelism. In the upper level, a program creates, in a lazy fashion, tasks to be scheduled on the different Streaming Multiprocessors (MP), as defined in the NVIDIA?s architecture. We have embedded inside GPU a well-known work stealing algorithm to dynamically balance the workload. At lower level, tasks exploit each Streaming Processor (SP) following a data-parallel approach. Preliminary comparisons on data-parallel iteration over vectors show that this approach is competitive on regular workload over the standard CUDA library Thrust, based on a static scheduling. Nevertheless, our approach outperforms Thrust-based scheduling on irregular workloads.

6.2. XKaapi on top of Multi-CPU Multi-GPU

Most recent HPC platforms have heterogeneous nodes composed of a combination of multi-core CPUs and accelerators, like GPUs. Programming such nodes is typically based on a combination of OpenMP and CUDA/OpenCL codes; scheduling relies on a static partitioning and cost model. We have experiment XKaapi runtime system for multi-CPU and multi-GPU architectures, which supports a data-flow task model and a locality-aware work stealing scheduler. The XKaapi enables task multi-implementation on CPU or GPU and multi-level parallelism with different grain sizes. We demonstrate performance results on two dense linear algebra kernels, matrix product (GEMM) and Cholesky factorization (POTRF), to evaluate XKaapi on a heterogeneous architecture composed of two hexa-core CPUs and eight NVIDIA Fermi GPUs. Our conclusion is two-fold: First, fine grained parallelism and online scheduling achieve performance results as good as static strategies, and in most cases outperform them. This is due to an improved work stealing strategy that includes locality information; to a very light implementation of the tasks in XKaapi; and to an optimized search for ready tasks. Next, our XKaapi Cholesky is highly efficient on multi-CPU/multi- GPU due to its multi-level parallelism. Using eight NVIDIA Fermi GPUs and four CPUs, we measure up to 2.43 TFlop/s on double precision matrix product and 1.79 TFlop/s on Cholesky factorization; and respectively 5.09 TFlop/s and 3.92 TFlop/s in single precision. This is the first time that such a performance is obtained with more than four GPUs.

6.3. Formalizing the concept of cooperation

We study how to optimize scheduling problems for a large number of objectives, when multiple users are competing for common resources, with some appropriate notion of fairness between users. Formalizing the concept of cooperation in relation with multi-objective optimization, we can refine the classical methods in combinatorial optimization (that usually optimize one centralized objective) by introducing extra features (adding more objectives or constraints). The PhD thesis of Daniel Cordeiro [2] proposed various ways for handling this problem: multi-organization scheduling and its relaxed variant, impact of selfishness. In the same context, we investigated the field of Game Theory through the existence of Nash equilibria in some situations.

6.4. Fault-tolerance for large parallel systems

This PhD thesis of Slim Bouguerra [1] studied fault-tolerance issues for large parallel systems. We revisited, via a formal proof, the old well-known result which states that the optimal policy for exponential failure law is to put the check-points at periodic moments. We proposed new algorithms to handle check-points for any law in the input and variable check-point costs (JPDC paper).

MYRIADS Project-Team

6. New Results

6.1. Autonomous Management of Virtualized Infrastructures

Participants: Amine Belhaj, Alexandra Carpen-Amarie, Roberto-Gioacchino Cascella, Stefania Costache, Djawida Dib, Florian Dudouet, Eugen Feller, Piyush Harsh, Rémy Garrigue, Filippo Gaudenzi, Ancuta Iordache, Yvon Jégou, Sajith Kalathingal, Christine Morin, Anne-Cécile Orgerie, Nikos Parlavantzas, Yann Radenac.

6.1.1. Application Deployment in Cloud Federations

Participants: Roberto-Gioacchino Cascella, Florian Dudouet, Piyush Harsh, Filippo Gaudenzi, Yvon Jégou, Christine Morin.

The move of users and organizations to Cloud computing will become possible when they will be able to exploit their own applications, applications and services provided by cloud providers as well as applications from third party providers in a trustful way on different cloud infrastructures. In the framework of the Contrail European project [17], we have designed and implemented the Virtual Execution Platform (VEP) service in charge of managing the whole life cycle of OVF distributed applications under Service Level Agreement rules on different infrastructure providers [43]. In 2012, we designed the CIMI inspired REST-API for VEP 2.0 with support for Constrained Execution Environment (CEE), advance reservation and scheduling service, and support for SLAs [40], [29], [32]. We integrated support for delegated certificates and provided test scripts to the Virtual Infrastructure Network (VIN) team. VEP 1.1 was slightly modified to integrate the usage control (Policy Enforcement Point (PEP)) solution developed by CNR. Work is in full progress to implement the CEE management interface and a complete web-based platform for all tasks.

6.1.2. Energy Management in IaaS Clouds: A Holistic Approach

Participants: Eugen Feller, Christine Morin.

Energy efficiency has now become one of the major design constraints for current and future cloud data center operators. One way to conserve energy is to transition idle servers into a lower power-state (e.g. suspend). Therefore, virtual machine (VM) placement and dynamic VM scheduling algorithms are proposed to facilitate the creation of idle times. However, these algorithms are rarely integrated in a holistic approach and experimentally evaluated in a realistic environment. We have designed overload and underload detection and mitigation algorithms and implemented them as well as a modified version of the Sercon existing consolidation algorithm [69] and power management algorithms and mechanisms in a novel holistic energy-efficient VM management framework for IaaS clouds called Snooze [25], [39]. In collaboration with David Margery and Cyril Rohr, we have conducted an extensive evaluation of the energy and performance implications of our system on 34 power-metered machines of the Grid'5000 experimentation testbed under dynamic web workloads. The results show that the energy saving mechanisms allow Snooze to dynamically scale data center energy consumption proportionally to the load, thus achieving substantial energy savings with only limited impact on application performance [26], [48]. Snooze has been released as an open source software since May 2012. It will be further developed and maintained as part of the Snooze ADT. This work has been carried out in the framework of Eugen Feller's PhD thesis [24], [8] funded by the ECO-GRAPPE ANR project.

6.1.3. A Case for Fully Decentralized Dynamic VM Consolidation in Clouds

Participants: Eugen Feller, Christine Morin.

One way to conserve energy in cloud data centers is to transition idle servers into a power saving state during periods of low utilization. Dynamic virtual machine (VM) consolidation (VMC) algorithms are proposed to create idle times by periodically repacking VMs on the least number of physical machines (PMs). Existing works mostly apply VMC on top of centralized, hierarchical, or ring-based system topologies, which result in poor scalability and/or packing efficiency with increasing number of PMs and VMs. We have proposed a novel fully decentralized dynamic VMC schema based on an unstructured peer-to-peer (P2P) network of PMs. The proposed schema is validated using three well known VMC algorithms: First-Fit Decreasing (FFD), Sercon, V-MAN, and a novel migration-cost aware ACO-based algorithm we have designed. Extensive experiments performed on the Grid'5000 testbed show that once integrated in our fully decentralized VMC schema, traditional VMC algorithms achieve a global packing efficiency very close to a centralized system. Moreover, the system remains scalable with increasing numbers of PMs and VMs. Finally, the migration-cost aware ACO-based algorithm outperforms FFD and Sercon in the number of released PMs and requires less migrations than FFD and V-MAN [23], [47]. This work has been done in the context of Armel Esnault's Master internship [57].

6.1.4. Market-Based Automatic Resource and Application management in the Cloud

Participants: Stefania Costache, Nikos Parlavantzas, Christine Morin.

Themis is a market-based Platform-as-a-Service system for private clouds. Themis dynamically shares resources between competing applications to ensure a fair resource utilization in terms of application priority and actual resource needs. Resources are allocated through a proportional-share auction while autonomous controllers apply elasticity rules to scale application demand according to resource availability and user priority. Themis provides users the flexibility to adapt controllers to their application types, and thus it can support diverse application types and performance goals. We have evaluated Themis through simulation and the obtained results demonstrated the effectiveness of the market-based mechanism[19], [20]. We have recently improved Themis in three ways. First, we extended the resource allocation algorithms to support multiple resources (CPU and memory) and to perform load-balancing between physical nodes while considering the migration cost. Second, we improved the management of applications. We added generic support for virtual cluster deployment, configuration and runtime management and also for application monitoring. Finally, we implemented several adaptation policies to scale elastically applications in term of number of provisioned virtual machines and in term of allocated CPU and memory per virtual machine. Themis is implemented in Python and uses OpenNebula for virtual machine operations. We used Themis to scale elastically two resource management frameworks (Torque and Condor) according to their current workload and also MPI scientific codes according to user-given deadlines. Themis has been deployed on Grid'5000 and also on EDF's testbed, HPSLAB. This work is carried out in the fraemwork of Stefania Costache's PhD thesis.

6.1.5. Autonomous PaaS-level resource management

Participants: Djawida Dib, Christine Morin, Nikos Parlavantzas.

PaaS providers host client applications on provider-owned resources or resources leased from public IaaS clouds. The providers have service-level agreements (SLAs) with their clients specifying application quality requirements and prices. A main concern for providers is sharing their private and leased resources among client applications in order to reduce incurred costs. We have proposed a PaaS architecture based on multiple elastic virtual clusters (VCs), each associated with a specific application type (e.g., batch, MapReduce). The VCs dynamically share the private resources using a decentralised allocation scheme and, when necessary, lease remote resources from public clouds. Resource allocation is guided by the SLAs of hosted applications and resource costs. We have implemented a prototype of this architecture that supports batch and MapReduce applications; the application SLAs constrain completion times and prices. The prototype is currently being evaluated on Grid'5000. This work is performed as part of Djawida Dib's thesis.

6.1.6. Elastic MapReduce on Top of Multiple Clouds

Participants: Ancuta Iordache, Yvon Jégou, Christine Morin, Nikos Parlavantzas.

We have worked on the design and implementation of Resilin. To the best of our knowledge Resilin is the first system which is capable of leveraging resources distributed across multiple potentially geographically distinct locations. Unlike the Amazon s proprietary Elastic Map Reduce (EMR) system, Resilin allows users to perform MapReduce computations across a wide range of resources from private, community, and public clouds such as Amazon EC2. Indeed, Resilin can be deployed on top of most of the open-source and commercial IaaS cloud management systems. Once deployed, Resilin takes care of provisioning Hadoop clusters and submitting MapReduce jobs thus allowing the users to focus on writing their MapReduce applications rather than managing cloud resources. In 2012 we designed and implemented a new version of Resilin based on a service-based architecture, which enables system recovery from errors and can be easily extended and maintained. Important functionalities were added to the system: scaling down the platform, deployment of data analysis systems (Apache Hive, Apache Pig). We have also started to work on the design of policies and mechanisms for the autonomous scaling of the virtual Hadoop clusters managed by Resilin. We perfomed an extensive experimental evaluation of Resilin on top of Nimbus and OpenNebula clouds deployed on multiple clusters of the Grid 5000 experimentation testbed. Our results show that Resilin enables the execution of MapReduce jobs across geographically distributed resources with only a limited impact on the jobs execution time, which is the result of intercloud network latencies [51], [31]. Resilin has been released as an open source software since September 2012. This work was carried out in the framework of the RMAC EIT ICT Labs activity.

6.1.7. Adaptation of the CooRM architecture into XtreemOS

Participants: Amine Belhaj, Rémy Garrigue, Yvon Jégou, Christine Morin, Yann Radenac.

In the framework of the COOP ANR project, we have mainly worked on the adaptation and on the implementation of the CooRM architecture (resulting from the work of the Avalon team at Inria Grenoble - Rhône Alpes in the context of the COOP project) into XtreemOS. The main results include a first version of the design of a decentralized version of CooRM, the modification of XtreemOS to support distributed applications (tested with OpenMPI and MPICH2), and the implementation of a launcher of moldable MPI applications using the modified XtreemOS API. A demonstration was presented to the COOP consortium in December 2012.

To get an operational prototype for evaluation purposes, we also had to fix many bugs in XtreemOS, revise its build chain, help clean the distribution package dependencies in collaboration with Rémy Garrigue (engineer from the ADT XtreemOS Easy), rewrite the code generator, help fix issues related to configuration commands in collaboration with Amine Belhaj (engineer from ADT XtreemOS Easy).

6.1.8. Extending a Grid with Virtual Resources Provisioned from IaaS Clouds

Participants: Amine Belhaj, Alexandra Carpen-Amarie, Rémy Garrigue, Sajith Kalathingal, Yvon Jégou, Christine Morin, Yann Radenac.

XtreemOS is a Grid operating system designed to facilitate the execution of grid applications by aggregating resources on multiple sites. XtreemOS provides virtual organization support and enables Grid users to run applications on the resources made available by their virtual organization. As the number of scientific applications that need access to Grid platforms increases, as well as their requirements in terms of processing power, the limited amount of resources that XtreemOS gathers from its virtual organizations may become a bottleneck. To address this limitation, we extended XtreemOS with the capability to acquire virtual resources from cloud service providers. To this end, we enable XtreemOS to provision and configure cloud resources both on behalf of a user and of a virtual organization. This can be done either on-demand, when a user specifically requires cloud resources, or in a dynamic fashion, when the local grid resources cannot comply with the application needs. Furthermore, we devised a selection mechanism for the cloud service providers, allowing users to rent resources from the providers that best match the requirements of their applications. We implemented our approach as a set of extension modules for XtreemOS and we evaluated the prototype in Grid'5000, using cloud resources provisioned from a private OpenNebula cloud. For this evaluation, we made a extensive use of tools developed jointly by Ascola and Myriads project-teams to easily manage large number of VMs on top of IaaS cloud management software (e.g. OpenNebula, Nimbus, OpenStack) deployed on the Grid'5000 platform. This work was carried out as part of the ANR Cloud project [60], [58] and an EIT ICT Labs activity.

6.1.9. Data Management Frameworks for Scientific Applications in Cloud Environments

Participants: Eugen Feller, Christine Morin.

During Eugen Feller's internship at LBNL, we have worked with Lavanya Ramakrishnan from the Advanced Computing for Science department on the evaluation of Hadoop MapReduce jobs in a virtualized environment. We have investigated the performance and power consumption of scientific MapReduce jobs executed in an environment with separated Hadoop compute and data nodes. This enables data sharing across multiple users and is key to support elastic MapReduce. Snooze cloud management stack was used to manage the VMs. Preliminary experimental results on top of Snooze demonstrate the feasibility of our approach.

6.1.10. Energy Consumption Models and Predictions for Large-scale Systems

Participant: Christine Morin.

We have collaborated with Taghrid Samak from the Advanced Computing for Science department at LBNL on the initial investigation of energy consumption models for Grid'5000 sites using Pig and Hadoop, and data from 6 months logs on 135 resources in the Lyon site. The initial results investigate time-series summarization for the entire dataset. For each resource the average power consumption is evaluated and compared with statistically estimated thresholds. A paper is under preparation.

6.1.11. Management of Large Data Sets

Participant: Christine Morin.

Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASAs satellites continuously generates data important to many scientific analyses. A dataprocessing pipeline that downloads the MODIS products, reprojects them on HPC systems or clouds and make them available to users through a web portal has been developed. In collaboration with Valerie Hendrix and Lavanya Ramakrishnan from the Advanced Computing for Science department at LBNL we have worked on providing community access to MODIS Satellite Reprojection and Reduction Pipeline and Data Sets. In a future version of the system, users will be able to reproject data on demand and/or run algorithms on the reprojected MODIS data such as an evapotranspiration calculation [30].

6.2. Dynamic Adaptation of Service-based Applications

Participants: Djawida Dib, Erwan Daubert, Guillaume Gauvrit, André Lage, Christine Morin, Nikos Parlavantzas, Jean-Louis Pazat, Chen Wang.

6.2.1. Adaptation for Service-Oriented Architectures

Participants: Erwan Daubert, Guillaume Gauvrit, André Lage, Nikos Parlavantzas, Jean-Louis Pazat, Chen Wang.

Service-Oriented Computing is a paradigm that is rapidly spreading in all application domains and all environments - grids, clusters of computers, mobile and pervasive platforms. The following works take place in the context of the S-CUBE European Network of Excellence.

6.2.1.1. Services adaptation in distributed and heterogeneous systems

Participants: Erwan Daubert, Guillaume Gauvrit, Jean-Louis Pazat.

We are still studying service adaptation in distributed and heterogeneous systems. This work covers different aspects such as structural, behavioral and environmental adaptation, distributed decision and planification of adaptation actions, adaptive allocation of resources for services. A framework called SAFDIS for "Self Adaptation For Distributed Services" has been defined and implemented. It is built as a set of services, providing functionalities useful to build an adaptation system. The analysis phase can take reactive as well as proactive decisions. This gives the ability to either react fast or to take decisions for the long term. This implies the ability to analyze the context with a variable depth of reasoning. Our implementation of the SAFDIS analysis phase also distributes and decentralizes its analysis process to spread the computational load and make the analysis process scalable. The planning phase seeks the set of actions (the plan) needed to adapt the system according to the strategy chosen by the analysis phase. It also schedules the selected actions to ensure a coherent and efficient execution of the adaptation. The planning topic is a well known subject in AI research works and many algorithms already exist in that field to produce efficient schedules. With our SAFDIS framework, the planning phase is able to reuse these algorithms. The resulting plan of actions can have actions that can be executed in parallel.

6.2.1.2. Quality Assurance for Distributed Services

Participants: André Lage, Nikos Parlavantzas, Jean-Louis Pazat.

In the context of the service-centric paradigm, we have designed and developed the Qu4DS (Quality Assurance for Distributed Services) system. Qu4DS is a cloud PaaS solution which fills the gap between SaaS service providers and IaaS infrastructures. Qu4DS provides automatic support for service execution management, aiming at increasing service providers' profits by reducing resource costs as well as fines owning to SLA violations. More specifically, Qu4DS dynamically acquires resources according to the customer demand, deploys service instances and implements QoS assurance mechanisms in order to prevent SLA violations. Qu4DS has been evaluated on Grid'5000 and shown to be effective in reducing service provider's costs [33]. This work has been done in the context of André Lage-Freitas' PhD thesis [10].

6.2.1.3. Self-configuration for Cloud Platforms

Participants: Jean-Louis Pazat, Chen Wang.

By definition, cloud computing offers an abstraction to manage various needs and concepts such as distributed software design, the deployment of such software on dynamic resources and the management of this kind of resources. Thus it is possible to reconfigure (adapt) according to some needs the software as well as the use of the resources. However these reconfigurations that are used on different layers may also have impacts on the others. Moreover these layers are independent and so are able to adapt themselves independently of the others. In our work, we propose to use some adaptation capabilities offered for example by the infrastructure (IaaS) that manages the resources to adapt the software (SaaS). We also propose to use planning algorithms to coordinate the adaptations between them to avoid inconsistency or inefficiency due to concurrent adaptations.

6.2.1.4. Dynamic Adaptation of Chemical services

Participants: Jean-Louis Pazat, Chen Wang.

We have proposed a QoS-aware middleware for dynamic service execution. In the context of dynamic execution, a workflow is defined by composing a set of abstract activities as place holders. Each activity is bound to a suitable partner service, which is selected at run-time from a set of functional equivalent candidates with different non-functional properties such as quality of service (QoS). The service selection process is modeled as a series of chemical reactions. This year, we have studied and implemented fragment replacement in workflows within this environment.

6.2.2. Multi-level Adaptation for Distributed Operating Systems

Participants: Djawida Dib, Christine Morin, Nikos Parlavantzas.

This work focused on enhancing distributed operating systems with the ability to continually adapt to their changing environments. Two challenges arise in this context: how to design the distributed operating system (OS) in order to facilitate dynamic adaptation, and how to ensure that OS-level adaptation does not conflict with application-level adaptation. This work proposed to address these challenges by (1) building the distributed OS as an assembly of adaptable services following the service-oriented architecture; and (2) using a common multi-level adaptation framework to adapt both the OS and the application layers in a coordinated way. To demonstrate the usefulness of the proposed architecture, the work focused on distributed shared memory services and provided examples and experimental results using the Kerrighed distributed OS. The work was performed as part of Djawida Dib's thesis [22].

6.3. A Chemical Approach for Autonomous Service Computing

Participants: Héctor Fernández, Marko Obrovac, Cédric Tedeschi.

6.3.1. Chemical Computing for the Simulation of Agile-Based Software Engineering

Participants: Héctor Fernández, Cédric Tedeschi.

In the framework of Héctor Fernández' internship at Vrije University, we applied the chemical programming model to simulate the behavior of a team developing software with Agile methods. Although an unexpected application, it has been the occasion to widen the range of applications and users of the software prototype developed during Héctor's thesis.

6.3.2. Scalable Atomic Capture of Molecules

Participants: Marko Obrovac, Cédric Tedeschi.

Capturing the reactants involved in a reaction constitutes one of the main challenges in the execution of chemical programs. Doing it at large scale is one of the essential barriers hindering the actual execution of chemical programs at large scale. While the problem resembles the classic resource allocation problem, it differs from it by different aspects. One of the main difference stands in the fact that the probability of a conflict varies during the course of execution. When the number of possible reactions is high, then there is no need for a complex conflict resolution scheme, as it would lead to a useless additional cost. In contrary, when this number drops, the probability of a conflict increases, and a systematic conflict resolution is mandatory to ensure at least one reaction will take place.

An adaptive protocol has been proposed, based on the dynamic combination of several strategies. Based on simulations, we have shown that, by dynamically switching from one strategy to another one, even by locally deciding which protocol to use, it is possible to combine the good properties of the strategies without suffering from their drawbacks [18].

The work was recently extended to take several rules into account. Rules have been defined to be able, not only to choose a strategy, but also to choose the rule to be executed, with the constant objective of maximizing the number of reactions executed in a given time.

6.3.3. DHT-based Runtime for the Chemical Programming Model

Participants: Marko Obrovac, Cédric Tedeschi.

The development of a distributed chemical machine entered its experimental phase with the development of a software prototype containing the following building blocks:

- A distributed hash table structures the network and allows any node to communicate with any other node in a logarithmic number of hops in this logical overlay.
- On top of the distributed hash table, a set of discovery mechanisms allows to find molecules needed in reactions, whatever their location is. These mechanisms are based on complex distribution and retrieval scheme borrowed from the P2P literature.
- The atomic capture protocol described before has been fully integrated in this framework.
- The discovery of molecules has been extended in order to detect the termination of the program and to be able to send the results of the computation back to the requester.

This software prototype has been deployed over the Grid'5000 platform [36].

OASIS Project-Team

6. New Results

6.1. Programming and Composition Models for Large-Scale Distributed Computing

6.1.1. Multi-active Objects

Participants: L. Henrio, F. Huet, A. Bourdin.

The active object programming model is particularly adapted to easily program distributed objects: it separates objects into several *activities*, each manipulated by a single thread, preventing data races. However, this programming model has its limitations in terms of expressiveness – risk of deadlocks – and of efficiency on multicore machines. We proposed to extend active objects with *local multi-threading*. We rely on declarative *annotations* for expressing potential concurrency between requests, allowing easy and high-level expression of concurrency. This year we realized the following:

- improvement on the model and its formalisation
- use of the new model in our CAN P2P network (see below); this was also the opportunity to improve our implementation.

This year, we also spent considerable efforts to publish this work; a conference paper is currently under review.

6.1.2. Events for Algorithmic skeletons

Participant: L. Henrio.

In the context of the SCADA associated team, we worked on the algorithmic skeleton programming model. The structured parallelism approach (skeletons) takes advantage of common patterns used in parallel and distributed applications. The skeleton paradigm separates concerns: the distribution aspect can be considered separately from the functional aspect of an application.

- This year we focused on the handling of events in algorithmic skeletons: adding the possibility for a skeleton to output an event should increase the control and monitoring capabilities of algorithmic skeletons. The ultimate goal is to improve autonomicity for algorithmic skeletons.

6.1.3. Behavioural models for Distributed Components

Participants: E. Madelaine, N. Gaspar, A. Savu, L. Henrio.

In the past [3], we defined the behavioural semantics of active objects and components. This year we extended this work to address group communications. On the practical side, this work contributes to the Vercors platform; the overall picture being to provide tools to the programmer for defining his application, including its behavioural specification. Then some generic properties like absence of deadlocks, but also application specific properties can be validated on the composed model using an existing model-checker. We mainly use the CADP model-checker, that also supports distributed generation of state-space. This year our main achievements are the following:

- We entirely formalised the specification of the behavioural model generation for component systems. This should provide us both a stronger formal background for our works in this area, and a specification for the automatic generation of behavioural models for our component systems.
- We additionally have put considerable efforts on the improvement of the Vercors platform and its integration with the Papyrus framework (see Section 5.2).

The formal work has been published as a research report [40]. A journal version is under submission. This work was done in collaboration with Rabéa Ameur-Boulifa from Télécom-Paristech.

In parallel with core developments of the behavioural specification environment, our collaborations led us to the study of the following application domain. In the context of the Spinnaker project, we are interested in developing a component-based distributed application to manage and monitor some pre-existing component-based distributed application - and hence, we called it The HyperManager. Our in-house component model (GCM) provides all the means to define, compose and dynamically reconfigure such applications. However, special care must be taken for this kind of undertaking. To this end, this year:

- We made the first steps towards a platform for the mechanized specification and verification, in the Coq Proof Assistant, of GCM applications. This work was published in [33], and is progressively being updated [1] to cope with behavioural specification, and to seamlessly combine deductive and model-checking techniques.
- We studied a real-life application scenario for our HyperManager prototype using distributed model-checking techniques in order to cope with the huge space state generated from reconfigurable applications.

6.1.4. Autonomic Monitoring and Management of Components

Participants: F. Baude, C. Ruz, B. Sauvan.

We have completed the design of a framework for autonomic monitoring and management of component-based applications. We have provided an implementation using GCM/ProActive taking advantage of the possibility of adding components in the membrane. For this purpose, we finalized the implementation of a factory which, from any GCM ADL description can instantiate the requested non functional components of a GCM application.

The framework for autonomic computing allows the designer to describe in a separate way each phase of the MAPE autonomic control loop (Monitoring, Analysis, Planning, and Execution), and to plug them or unplug them dynamically. We have demonstrated how such a control loop can be relevant to drive the dynamic reconfiguration of services part of a SOA application, considering as in the SCA standard, that services are components [15].

Our objective now is to exemplify such autonomic and structured approach in the management of any distributed middleware or application, e.g. in the Spinnaker industrial context.

6.1.5. Optimization of data transfer in SOA and EDA models

Participants: I. Alshabani, F. Baude, L. Pellegrino, B. Sauvan, Q. Zagarese.

Traditional client-server interactions rely upon method invocations with copy of the parameters. This can be useless in particular if the receiver does not effectively uses them. On the contrary, copying and transferring parameters lazily so to allow the receiver to proceed even without all of them is a meaningful idea that we proved to be effective for active objects in the past [56]. This idea wasn't so far realized in the context of the web services technology, the most popular one used today for client-server SOAP-based interactions.

- To such an aim, we contributed to the offloading of objects representing parameters of the web service Java Apache CXF API [29]. It is innovative notably in the way the offloading of parameters for on-demand access can be delegated from services to services, which resembles the concept of first-class futures.
- Relying upon such an effective approach, we have applied a similar idea of “lazy copying and transfer” to the data parts of events in the context of event-driven architecture applications [28]. The middleware dynamically off-loads data (generally of huge size) attached to an event, according to some user-level policy expressed as annotation in the Java code at the subscriber side. The event itself, without its attachments, gets forwarded into the publish/subscribe brokering system (in our case, the event cloud middleware, that is the subject of section 6.2.1) and its attachements are transferred to the subscriber only on-demand. Compared to some existing propositions geared towards a data centric publish-subscribe pattern (e.g. the DDS OMG standard), ours is more user-friendly as it does not require the user code to explicitly program when to get the data attached to notified events.

[1] http://www-sop.inria.fr/members/Nuno.Gaspar/Mefresa.php

Overall, this work opens the way towards a strong convergence between service oriented and event-driven technologies.

6.1.6. Multi-layer component architectures

Participant: O. Dalle.

Since a few years, we have been investigating the decomposition of a simulation application into multiple layers corresponding to the various concerns commonly found in a simulation: in addition to the various modeling domains that may be found in a single simulation application (e.g. telecommunications networks, road-networks, power-grids, and so on), a typical simulation includes various orthogonal concerns such as system modelling, simulation scenario, instrumentation and observation, distribution, and so on. This large number of concerns has put in light some limits of the traditional hierarchical component-based architectures and their associated ADL, as found in the FCM and GCM. In order order to cope with these limitations, we started a new component architecture model called Binding Layers centered on the binding rather than the component, with no hierarchy but advanced layering capabilities, and offering advanced support for dynamic structures[32].

6.2. Middleware for Grid and Cloud computing

6.2.1. Publish-Subscribe in Distributed Environments

Participants: F. Baude, F. Huet, F. Bongiovanni, L. Pellegrino, B. Sauvan, I. Alshabani, A. Bourdin, M. Antoine, A. Alshabani.

In the context of the SOA4ALL FP7-IP project, we designed and implemented a hierarchical Semantic Space infrastructure based on Structured Overlay Networks (SONS) [62], [63]. It originally aimed at the storage and the retrieval of the semantic description of services at the Web scale [57]. This infrastructure combines the strengths of both the P2P paradigm at the architectural level and the Resource Description Framework (RDF) data model at the knowledge representation level. The achievements of this year are the following:

- In the context of the FP7 Strep PLAY and French ANR SocEDA research projects, we have been extending the aforementioned work with a content-based Publish/Subscribe abstraction in order to support asynchronous queries for RDF-based events in large scale settings, which raises some interesting challenges [26]. The goal is to build a platform for large scale distributed reasoning[25]. Such an integrated working platform [39], [38] has been presented in two tutorials [27], [54].
- We have also investigated the Publish/Subscribe paradigm in the MapReduce programming model. We have proposed the concept of continuous job which allows MapReduce jobs to be re-executed when new data are added to the system. To maintain the correctness of the execution, we have introduced the notion of carried data, i.e. data which are kept between subsequent executions. An implementation has been written on top of Hadoop and a paper submitted.

6.2.2. Distributed algorithms for CAN-like P2P networks

Participants: L. Henrio, F. Bongiovanni, F. Huet.

The nature of some large-scale applications, such as content delivery systems or publish/subscribe systems, built on top of SONs, demands application-level dissemination primitives which do not overwhelm the overlay, i.e. efficient, and which are also reliable. Building such communication primitives in a reliable manner on top of such networks would increase the confidence regarding their behavior prior to deploying them in real settings. In order to come up with real efficient primitives, we take advantage of the underlying geometric topology of the overlay network and we also model the way peers communicate with one another. Our objective is to design and prove an efficient and reliable broadcast algorithm for CAN-like P2P networks. To this aim, in 2012 we:

- Improved the formalisation in Isabelle/HOL of a CAN-like P2P system, devised formalised tools to reason on CAN topologies, and on communication protocols on top of CANs. We designed and proved the efficiency of a first naive algorithm.
- Sketched on paper the proof of completeness and efficiency for the algorithm we designed and implemented last year.

Part of this work was done in the PhD thesis of F. Bongiovanni [10]

We are also investigating the new algorithms to efficiently build a SONs in the presence of existing data. Most of the work on SONs assume that new peers joining the network will arrive without data or fail to take into account the cost of distributing these data. Indeed, depending on the key subspace given to the new peer, some or all its data will have to be distributed in the network. In 2012:

- We proposed a first version of new join algorithms which try to allocate key sub-spaces to peers so that the amount of data that needs to be moved is minimal. An expected benefit of this work is that it should allow for fast and efficient reconstruction of a SON in case of a crash, without having to use distributed snaphshots.

6.2.3. Network Aware Cloud Computing

Participants: S. Malik, F. Huet.

We have worked on the Resource Aware Cloud Computing project. Its primary purpose is to address different issues which can help the scheduler to make more efficient scheduling decisions. These issues are related to the resource characteristics.

- We introduce a framework, which increases the performance of the application and ensures high level of reliability during the scheduling of application onto the cloud. It is a cloud scheduler module named as Resource Aware Cloud Scheduling (RACS) module. It helps the scheduler in making the scheduling decisions on the basis of different characteristics of cloud resources. These characteristics are reliability, network latency, and monetary cost. RACS consists of multiple sub modules, which are responsible for their corresponding tasks. In RACS, we have done the implementation for the different issues.
- We worked on a model for the reliability assessment of the cloud's computing nodes. This reliability assessment mechanism helps to do the scheduling on cloud infrastructure and perform fault tolerance on the basis of the reliability values acquired during reliability assessment. The model has different algorithms for different types of applications. Thus it has multiple reliability values for each computing node. For real time applications, the model has time based reliability assessment algorithms.

This work is part of S. Malik's PhD thesis [12]

6.2.4. Testbed Designs from Experimenters Requirements

Participant: F. Hermenier.

The physical design of the Emulab facility, and many other testbeds like it, has been based on the facility operators' expectations regarding user needs and behavior. If operators' assumptions are incorrect, the resulting facility can exhibit inefficient use patterns and sub-optimal resource allocation.

- We have collaborated with Robert Ricci from the University of Utah on the study of the Utah' Emulab facility to provide better testbed designs. Our study gained insight into the needs and behaviors of networking researchers by analyzing more than 500,000 topologies from 13,000 experiments submitted to Emulab.
- Using this dataset, we re-visited the assumptions that went into the physical design of the Emulab facility and considered improvements to it. Through extensive simulations with real workloads, we evaluated alternative testbeds designs for their ability to improve testbed utilization and reduce hardware costs.

The results have been published to TridentCom [22], the reference conference related to testbeds and research infrastructures, and the article received the best paper award.

6.2.5. Energy Efficient Virtual Machines Placement in Data Centers

Participant: F. Hermenier.

Data centres are powerful ICT facilities which constantly evolve in size, complexity, and energy consumption. At the same time, tenants' and operators' requirements become more and more complex. The data centre operators may target different energy-related objectives while the workload volatility may alter the data centre capacity at supporting load spikes. Finally, clients of data centres are looking for dependable infrastructures that can comply with their SLA requirements.

To stay attractive, a data centre should then support these expectations. These constraints are however very specific to each of the tenants but also to the infrastructure. They also cover a large range of concerns (hardware requirements, performance, security ...) that are continuously evolving according to new trends and new technologies. Existing solutions are however ad-hoc and can not be updated easily to fit the data centres and the workload specificities.

We proposed a flexible energy-aware framework to address the multiple facets of an energy-aware consolidation of VMs in a cloud data centre.[21] This framework extended BtrPlace to make it able to address specific energy concerns. We integrated a fine grain energy model reducing either gas emissions or power consumption. We also proposed constraints to control the aggressiveness of these objectives to let the data centre reactive when a load spike occurs. We finally proposed various constraints to satisfy the hardware and the resource requirements of the tenants. The evaluation on a testbed running an industrial workload validated the practical benefits provided by the usage of our framework.

6.2.6. GPU-based High Performance Cloud Computing

Participants: M. Benguigui, F. Baude, F. Huet.

To address HPC, GPU devices are now considered as unavoidable cheap, energy efficient and very efficient alternative computing units. The barrier to handle such devices is the programming model: it is both very fine grained and synchronous.

Our long term goal is to devise some generic solutions in order to incorporate GPU-specific code whenever relevant into a parallel and distributed computation. The first step towards this objective is to gain some insight on how to efficiently program a non trivial but well known algorithm. We selected the American basked option pricing non embarrassingly parallel problem that was previously parallelized and distributed using ProActive master-slave approach [60], achieving an almost linear speedup and good performances (64 CPUs based computation allowed us to solve the problem in about 8 hours). The same algorithm has been reorganized for running on **a single GPU** [17] and achieved the same option pricing computation in about 9 hours. The current work is to succeed to take advantage of GPUs, even if non homogeneous, hired from a Cloud or a federation of clouds at once, orchestrated by an active object acting as a GPU task delegator. The goal is to drastically lower the overall computation time for such highly time consuming stochastic simulation problems.

6.3. Large-scale Simulation Platform: Techniques and methodologies

Participants: O. Dalle, E. Mancini.

In the domain of simulation techniques and methodologies, this year, we conducted research in the three following areas:

Simulation in the Cloud In recent years, numerous applications have been deployed into mobile devices. However, until now, there have been no attempts to run simulations on handheld devices. In the context of the DISSIMINET Associate Team, we work in collaboration with our partners at the Carleton University to investigate different architectures for running and managing simulations on handheld devices, and putting the simulation services in the Cloud[24]. We propose a hybrid simulation and visualization approach, where a dedicated mobile application is running on the client side and the RISE simulation server is hosted in the Cloud.

Simulation Methodology In the context of the ANR INFRA SONGS project, we are involved (as coordinators) in a Work-Package called "Open Science" whose aim is to investigate and contribute means to ensure the long term visibility and reproductibility of simulation results obtained using the SimGrid simulation platform. Our preliminary work in this direction consisted in identifying the issues, trends and potential solutions to ensure the long-term reproducibility of simulations[16].

Peer-to-peer Simulation In order to evaluate the performance and estimate the resource usage of peer-to-peer backup systems, it is important to analyze the time they spend in storing, retrieving and keeping the redundancy of the stored files. The analysis of such systems is difficult due to the random behavior of the peers and the variations of network conditions. In the context of the ANR USS-SIMGRID and INFRA-SONGS projects, we investigated means for reproducing such varying conditions in a controlled way. We worked on the design of a general simulation meta-model for peer-to-peer backup systems and a tool-chain, based on SimGrid, to help in their analysis[20]. We validated the meta-model and tool-chain through the analysis of a common scenario, and verified that they can be used, for example, for retrieving the relations between the storage size, the saved data fragment sizes and the induced network workload. We also started to investigate a new simulation technique for very-large scale distributed simulation of peer-to-peer systems based on the decomposition of a simulation into many micro-simulation steps[31] in order to optimize the overlap between communications and computations.

PHOENIX Project-Team

6. New Results

6.1. Design-driven Testing by simulation

Previously, we have introduced a paradigm-oriented development approach that revolves around a conceptual framework concretized by a design language [26]. A design description is used to generate high-level programming support, to perform a range of verifications, and to abstract over underlying technologies.

This approach is illustrated with the Sense-Compute-Control (SCC) paradigm [48], where an SCC software system gathers information about an environment via sensors (whether hardware or software) and issues orders to impact the environment via actuators. The SCC paradigm has a wide spectrum of applicability; we have used it successfully in the domains of home/building automation, multimedia, avionics and networking.

SCC systems involve both software concerns, like any software system, and integration concerns, for the constituent networked entities forming the environment of the SCC-loop. This situation is problematic for testing because it requires acquiring, testing and interfacing a variety of software and hardware entities. This process can rapidly become costly and time-consuming when the target environment involves many entities.

We have developed a simulation approach and a tool named *DiaSim* that leverage the DiaSpec description of an environment [15]. This description is used to generate both a programming framework to develop the simulation logic and an emulation layer to execute applications. The generic nature of our approach has been illustrated by leveraging two different simulation tools, namely, Siafu for 2-D rendering of home/building spaces, and FlightGear for avionics.

To fuel the simulation of an environment with accurate stimuli, we need to model real systems, including natural phenomena (*e.g.,* heat transfer in a building) or mechanical systems (*e.g.,* aircraft models). These physical models are typically defined as continuous systems using differential equations. To facilitate the reuse of off-the-shelf physical models, we have used a DSL named Acumen for describing differential equations. Acumen continuous models are coupled with the DiaSim discrete simulator, forming a hybrid system fueled by accurate stimulus producers [18].

These major accomplishments were conducted by Julien Bruneau, in the context of his PhD studies [11].

6.2. Design-driven Development of Dependable Software Systems

Dependability of a system is the ability to avoid service failures that are more frequent and more severe than is acceptable [22]. This generic concept includes attributes such as availability, integrity and reliability. Dependable systems are now pervasive in a range of domains (*e.g.,* railway, avionics, automotive) and require a certification process. The main goal of certification is to demonstrate that a system is conform to its *high-level requirements*, resulting from functional and safety analyses.

Software plays an increasingly important role in dependable systems; software development is thus required to be certified. In particular, the stakeholders have to pay attention to the coherence of the functional and non-functional aspects of an application to demonstrate the conformance of the software with the high-level requirements. Non-functional aspects of a system refer to constraints on the manner in which this system implements and delivers its functionality (*e.g.,* performance, reliability, security) [48].
Coherence. Because functional and non-functional aspects are inherently coupled, ensuring their coherence is critical to avoid unpredicted failures [39]. For example, fault-tolerance mechanisms may significantly deteriorate the application performance. Generally, this kind of issues are detected at the late stages of the development process, increasing the development cost of applications [21].
Conformance. Ensuring that an application is in conformance with its high-level requirements is typically done by tracing their propagation across the development stages. In practice, this process is human-intensive and error prone because it is performed manually [37].

Certifying a development process requires a variety of activities. In industry, the usual procedures involve holding peer review sessions for coherence verification, and writing traceability documents for conformance certification. In this context, *design-driven development* approaches are of paramount importance because the design drives the development of the application and provides a basis for tracing requirements [53]. However, because most existing approaches are general purpose, their guidance is limited, causing inconsistencies to be introduced in the design and along the development process. This situation calls for an integrated development process centered around a conceptual framework that allows to guide the certification process in a systematic manner. In response to this situation, we proposed a design-driven development methodology, named DIASUITE [2], which is dedicated to the *Sense/Compute/Control (SCC) paradigm* [48]. As demonstrated by Shaw, the use of a specific paradigm provides a conceptual framework, leading to a more disciplined engineering process and guiding the verification process [47]. An SCC application is one that interacts with a physical environment. Such applications are typical of domains such as home/building automation, robotics and avionics.

In this work, we have shown the benefits of DIASUITE for the development of dependable SCC applications. This approach is applied to a realistic case study in the avionics domain, in the context of two non-functional aspects, namely time-related performance and reliability. The DIASUITE design language, named DIASPEC , offers declarations covering both functional and non-functional dimensions of an SCC application [2], [9] [32]. However, so far, the DIASUITE methodology has only been used to study each dimension in isolation, leaving open the problems of coherence and conformance when considering multiple dimensions. This work integrates all these dimensions, enabling the generation of validation support. More precisely, this work makes the following contributions:
Design coherence over functional and non-functional dimensions. We use the DIASPEC language to describe both functional and non-functional aspects of an application and apply this approach to a realistic case study. A DIASPEC description is verified at design time for coherence of its declarations. This verification is performed with respect to a formal model generated from a DIASPEC description.
Design conformance through the development process. At design time, we provide verification support to check the conformance between the specification and the formalized form of the high-level requirements. At implementation time, we guarantee the conformance between the application code and the previously verified requirements. This process is automatically done by leveraging the generative approach of DIASUITE . As some of the high-level requirements cannot be ensured at design time (*e.g.,* time-related performance), we provide further testing support to validate the implementation with respect to these remaining requirements. This support leverages a realistic flight simulator, namely FlightGear [44].
Validation in avionics. We validate our approach by developing a realistic case study in avionics. Following the DIASUITE methodology, we have developed an aircraft flight guidance system and tested it on FlightGear. Additionally, we have duplicated this case study in the context of a commercial drone system, namely Parrot AR.Drone. [2]

These accomplishments were conducted by Julien Bruneau, Quentin Enard and Stéphanie Gatti, in the context of their PhD studies. This work will be published at the International Conference on Pervasive and Embedded Computing and Computation Systems (PECCS'13).

6.3. Putting DiaSuite to Work

A continuing concern of the Phoenix research group is to put our work into practice by tackling realistic applications. We have validated DiaSuite on a variety of applications in areas including telecommunications, pervasive computing, and avionics.

Our expertise in smart home and building, combined with the maturity of DiaSuite, have given rise to the development of a dedicated instance of our technology called DiaSuiteBox. This instance is destined for technology transfer.

[2] http://ardrone.parrot.com

6.3.1. Applying DiaSuite to a Variety of Areas.

Let us examine the application of DiaSuite to two key areas: pervasive computing and avionics. In each case, demonstrations and posters have been presented to researchers and industrial partners [24], [25], [23], [35]

Smart Homes. Despite much progress, developing a pervasive computing application remains a challenge because of a lack of conceptual frameworks and supporting tools. This challenge involves coping with heterogeneous entities, overcoming the intricacies of distributed systems technologies, working out an architecture for the application, encoding it in a program, writing specific code to test the application, and finally deploying it.

At the beginning of this evaluation period, our research group was mainly interested in orchestrating applications in the telecommunications domain, leveraging new opportunities created by the emergence of Voice over IP (mainly based on SIP). Concurrently, a myriad of objects became networked, prompting a need to expand the scope of telecommunications beyond human-human interaction.

Two main industrial collaborations were instrumental to explore the scope of this evolution and to validate the Diasuite approach with realistic case studies. First, we collaborated with a French telecommunications company, in a two-year project named HomeSIP, to study the convergence between VoIP and networked objects in the context of home automation. During this project, we developed a range of applications, including remote appliance control through phone keypad, TV recording via SMS, and dynamic entry phone systems. Second, we contributed to a two-year project named SmartImmo, which gathered major French companies in the area of building construction, installation, and management. The goal of this project was to create a service infrastructure for building automation. SmartImmo gave us the opportunity to elaborate realistic building automation scenarios (*e.g.,* parking lot management, meeting room reservation, energy monitoring).

Our work on applying DiaSuite to the pervasive computing domain has leveraged key contributions by two PhD students of Phoenix, namely, Wilfried Jouve [36] and Nicolas Palix [42]. They both defended at the beginning of this evaluation period.

Avionics. Safety-critical applications have to fulfill stringent requirements, both functional and non-functional. These requirements have to be coherent with each other and must be preserved throughout the software development process. In this context, a design-driven development approach can play a critical role. However existing design-driven development approaches are often general purpose, providing little, if any, conceptual framework to guide the development. Previously, we explained how the DiaSuite approach was enriched with non-functional declarations such as QoS and error handing.

To validate the interest of DiaSuite for safety-critical applications, several avionics case studies have been realized in the context of a collaboration with Thales, a French airborne systems company. One case study was a flight guidance application; it is in charge of the plane navigation and is under the supervision of the pilot. For example, if the pilot specifies a heading to follow, the application compares it to the current heading, sensed by devices such as the Inertial Reference Unit, and maneuvers ailerons accordingly. To test this application, we have used the DiaSim tool coupled with the FlightGear simulator. A flight guidance application has also been developed for a commercial drone platform. The goal of this application was to make the drone autonomous by following a flight plan similar to the one in avionics.

This simulation work has been presented in the thesis of Julien Bruneau [11]. Non-functional concerns addressing error handling and QoS will be presented in two forthcoming dissertations.

6.3.2. DiaSuiteBox: an Ongoing Technology-Transfer Project.

The DiaSuiteBox platform runs an open-ended set of applications, leveraging a range of appliances and web services. Our solution consists of a dedicated development environment, an application store, and a lightweight runtime platform. This solution is based on DiaSuite.

DiaSuiteBox consists of three main components:

- A tool-based environment is dedicated to the development of applications, orchestrating networked entities. This environment leverages DiaSpec, its compiler and an Eclipse plugin.
- An application store is composed of two servers: (1) a server verifies and packages submitted applications of developers prior to making them available to users and (2) another server enables users to browse, select and install applications.
- An execution environment runs end-user applications and allows to manage and configure all aspects of a smart space. This environment can either be deployed on low-resource computing platform (*e.g.,* Plug-PC, set-top-box) at the end-user's home or in the Cloud, coupled with a gateway for controlling equipments on the end-user's side.

Thanks to the application store and a developer community, the platform should provide users with a stream of innovative applications. During the submission process, an application is automatically analyzed and checked in order to be certified. [3] The user is ensured that the behavior of its applications is innocuous and conform to their description. DiaSuiteBox supports several technology standards like UPnP, Bluetooth, USB...This platform can be easily extended by plugging appliances directly on the hardware platform or by connecting devices on the local network.

[3] This certification process is preliminary in the current version of DiaSuiteBox.

PLANETE Project-Team

6. New Results

6.1. Towards Data-Centric Networking

Participants: Chadi Barakat, Damien Saucez, Jonathan Detchart, Mohamed Ali Kaafar, Ferdaouss Mattoussi, Marc Mendonca, Xuan-Nam Nguyen, Vincent Roca, Thierry Turletti.

- **DTN**

  Delay Tolerant Networks (DTNs) stand for wireless networks where disconnections may occur frequently. In order to achieve data delivery in such challenging environments, researchers have proposed the use of store-carry-and-forward protocols: there, a node may store a message in its buffer and carry it along for long periods of time, until an appropriate forwarding opportunity arises. Multiple message replicas are often propagated to increase delivery probability. This combination of long-term storage and replication imposes a high storage and bandwidth overhead. Thus, efficient scheduling and drop policies are necessary to: *(i)* decide on the order by which messages should be replicated when contact durations are limited, and *(ii)* which messages should be discarded when nodes' buffers operate close to their capacity.

  We worked on an optimal scheduling and drop policy that can optimize different performance metrics, such as the average delivery rate and the average delivery delay. First, we derived an optimal policy using global knowledge about the network, then we introduced a distributed algorithm that collects statistics about network history and uses appropriate estimators for the global knowledge required by the optimal policy, in practice. At the end, we are able to associate to each message inside the network a utility value that can be calculated locally, and that allows to compare it to other messages upon scheduling and buffer congestion. Our solution called HBSD (History Based Scheduling and Drop) integrates methods to reduce the overhead of the history-collection plane and to adapt to network conditions. The first version of HBSD and the theory behind have been published in 2008. A recent paper [27] provides an extension to a heterogenous mobility scenario in addition to refinements to the history collection algorithm. An implementation is proposed for the DTN2 architecture as an external router and experiments have been carried out by both real trace driven simulations and experiments over the SCORPION testbed at the University of California Santa Cruz. We refer to the web page of HBSD for more details http://planete.inria.fr/HBSD_DTN2/.

  HBSD in its current version is for point-to-point communications. Another interesting schema is to consider one-to-many communications, where requesters for content express their interests to the network, which looks for the content on their behalf and delivers it back to them. Along the main ideas of HBSD, we worked on a content optimal-delivery algorithm, CODA, that distributes content to multiple receivers over a DTN. CODA assigns a utility to each content item published in the network; this value gauges the contribution of a single content replica to the network's overall delivery-rate. CODA performs buffer management by first calculating the delivery-rate utility of each cached content-replica and then discarding the least-useful item. When an application requests content, the node supporting the application will look for the content in its cache. It will immediately deliver it to the application if the content is stored in memory. In case the request cannot be satisfied immediately, the node will store the pending request in a table. When the node meets another device, it will send the list of all pending requests to its peer; the peer device will try to satisfy this list by sending the requester all the matching content stored in its own buffer. A meeting between a pair of devices might not last long enough for all requested content to be sent. We address this problem by sequencing transmissions of data in order of decreasing delivery-rate utility. A content item with few replicas in the network has a high delivery rate utility; these items must be transmitted first to avoid degrading the content delivery-rate metric. The node delivers the requested content to the application as soon as it receives it in its buffer. We implement CODA over the CCNx protocol, which provides the basic tools for requesting, storing, and forwarding content. Detailed information on CODA and the implementation work carried out herein can be found in [76].

- **Naming and Routing in Content Centric Networks**

  Content distribution prevails in todays Internet and content oriented networking proposes to access data directly by their content name instead of their location, changing so the way routing must be conceived. We proposed a routing mechanism that faces the new challenge of interconnecting content-oriented networks. Our solution relies on a naming resolution infrastructure that provides the binding between the content name and the content networks that can provide it. Content-oriented messages are sent encaspulated in IP packets between the content-oriented networks. In order to allow scalability and policy management, as well as traffic popularity independence, binding requests are always transmitted to the content owner. The content owner can then dynamically learn the caches in the network and adapt its binding to leverage the cache use.

  The work done so far is related to routing between content-oriented networks. We are starting an activity on how to provide routing inside a content network. To that aim, we are investigating on the one hand probabilistic routing and, on the other hand, deterministic routing and possible extension to Bellman-Ford techniques. In addition to routing, we are investigating the problem of congestion in content-oriented networks. Indeed, in this new paradigm, congestion must be controlled on a per-hop basis, as opposed to the end-to-end congestion control that prevails today. We think that we can combine routing and congestion control to optimize resource consumption. Finally, we are studying the implications of using CCN from an economical perspective. See [100] for more details.

- **On the fairness of CCN**

  Content-centric networking (CCN) is a new paradigm to better handle contents in the future Internet. Under the assumption that CCN networks will deploy a similar congestion control mechanism than in today's TCP/IP (i.e., AIMD), we built an analytical model of the bandwidth sharing in CCN based on the "square-root formula of TCP". With this model we can compare CCN download performance to what users get today. We consider different factors such as the way CCN routers are deployed, the popularity of contents, or the capacity of links and observe that when AIMD is used in a CCN network less popular content throughput is massively penalised whilst the individual gain for popular content is negligible. Finally, the main advantage of using CCN is the decrease of load at the server side. Our observations advocate the necessity to clearly define the notion of fairness in CCN and to design a proper congestion control to avoid less popular contents to become hardly accessible in tomorrow's Internet.

  Our results [75] clearly point to a fairness issue if AIMD is used with CCN. Indeed, combining blindly AIMD and CCN can severely worsen the download throughput of less popular contents with respect to the today's Internet due to subtle interactions with in-network caching strategies. The way cache memories are distributed within chain topologies has been investigated too, showing that for small and heterogeneous cache spaces, placing the biggest caches close to clients improves performance due to a smaller RTT on average. On the other hand, CCN can significantly reduce the load at the server side independently of the cache allocation strategy. Our findings advocate the urge of clearly defining the notion of fairness in CCN and designing congestion control algorithms able to limit the unfairness observed between contents of different popularities. The work is currently used within the IRTF ICNRG research group in order to motivate and define an appropriate congestion control mechanism for information centric networks like CCN. Moreover, we are currently validating the analytical results with an implementation of CCN where we can evaluate how much our model deviates from the reality when contents are of various size or small. The implementation will also be a support to test different congestion control mechanism.

- **CCN to enable profitable collaborative OTT services**

  The ubiquity of broadband Internet and the proliferation of connected devices like laptops, tablets, or TV result in a high demand of multimedia content such as high definition video on demand (VOD) for which the Internet has been poorly designed with the Internet Protocol (IP). Information-Centric Networking and more precisely Content Centric Networking (CCN) overtake the limitation of IP by considering content as the essential element of the network instead of the topology. CCN and its content caching capabilities is particularly adapted to Over-The-Top (OTT) services like Netflix, Hulu, Xbox Live, or YouTube that distribute high-definition multimedia content to millions of consumers, independently of their location. However, bringing content as the most important component of the network implies fundamental changes in the Internet and the transition to a fully CCN Internet might take a long time. Despite this transition period where CCN and IP will co-exist, we have shown that OTT service providers and consumers have strong incentives for migrating to CCN. We also propose a transition mechanism based on the Locator/Identifier Separation Protocol (LISP) [28] that allows the provider to track the demands from its consumers even though they do not download the contents from another consumers instead of the producer itself.

  CCN, compared to IP, provides better security and performance. This last point is very interesting for OTT service providers that deliver multimedia content where performance is a key factor for the adoption of the service by consumers. With CCN, the content can be retrieved from the caches in the different CCN islands, instead of always being delivered by the content publisher. As a result, content retrieval is faster for the consumer and the operational cost of the publisher is reduced. Moreover, as the content is cached by the consumers and because the consumer can provide the content to other consumers, the overall performance increases with the number of consumers instead of decreasing as it is the case in IP today where the content is delivered by the hosting server. This property is particularly interesting because it dampens the effect of flash crowds which are normally very costly for OTT service providers as they have to provision their servers and networks to support them. Using CCN with caching at the consumers has then a direct impact on the profit earned by the OTT service provider as its costs are reduced. However, to benefit from the caching capabilities of consumers, the producer must propose real incentives to its consumers to *collaborate* and cache the content. To understand how incentives can be provided, it is necessary to remember that content in OTT is provided either freely to the consumer or in exchange of a fee. When the content is provided freely, the incomes for the publisher are ensured by advertisements dispersed in the content (e.g., banner, commercial interruptions...). A consumer has incentives to collaborate with the system if it receives some sort of discount, expressed in advertisement reduction or fee reduction. On the one hand, the discount has a cost for the publisher as its revenues will be reduced. On the other hand, the collaboration from its consumers reduces its operational costs. Hence, the publisher must determine the optimal discount, such that it maximises its profit. The situation for the consumer is the exact opposite: its costs are increasing because it is providing content to other consumers but its revenues also increase as it receives a discount on its expenses. We have determined the conditions to respect when deploying OTT with loosely collaborative consumers [99]. We currently refine the results using game theory.

- **Software-Defined Networking in Heterogeneous Networked Environments**

Software-Defined Networking (SDN) has been proposed as a way to facilitate network evolution by allowing networks and their infrastructure to be programmable. In the context of the COMMUNITY associated team with University of California Santa Cruz (see URL http://inrg.cse.ucsc.edu/community/) , we are studying the potential of SDN to facilitate the deployment and management of new architectures and services in heterogeneous environments. In particular, we focus on the fundamental issues related to enabling SDN in infrastructure-less/decentralized networked environments and we use OpenFlow as our target SDN platform. Our plan is to develop a hybrid SDN framework that strikes a balance between a completely decentralized approach like Active Networking and a centralized one such as OpenFlow~[58].

We are also currently evaluating the efficiency of SDN for optimizing caching in content-centric networks. CCN advocates in-network caching, i.e., to cache contents on the path from content providers to requesters. Akthough this on-path caching achieves good overall performance, we have shown that this strategy is far from being the optimal inside a domain. On this purpose, we proposed the notion of off-path caching by allowing deflection of the most popular traffic off the optimal path towards off-path caches available across the domain[100]. Off-path caching improves the global hit ratio and permits to reduce the peering links' bandwidth usage. We are now investigating whether SDN functionalities can be used to implement this optimal caching technique, in particular to identify of the most popular contents, and to configure deflection mechanisms within routers~[94].

- **Application-Level Forward Error Correction Codes (AL-FEC) and their Applications to Broadcast/Multicast Systems**

With the advent of broadcast/multicast systems (e.g., 3GPP MBMS services), large scale content broadcasting is becoming a key technology. This type of data distribution scheme largely relies on the use of Application Level Forward Error Correction codes (AL-FEC), not only to recover from erasures but also to improve the content broadcasting scheme itself (e.g., with FLUTE/ALC).

Our LDPC-Staircase codes, that offer a good balance in terms of performance, have been included as the primary AL-FEC solution for ISDB-Tmm (Integrated Services Digital Broadcasting, Terrestrial Mobile Multimedia), a Japanese standard for digital television (DTV) and digital radio, with a commercial service that started in April 2012. This is the first adoption of these codes in an international standard. These codes, along with our FLUTE/ALC software, are now part of the server and terminal protocol stack: http://www.rapidtvnews.com/index.php/2012041721327/ntt-data-mse-and-expways-joint-solution-powers-japanese-mobile-tv-service.html.

This success has been made possible, on the one hand, by major efforts in terms of standardization within IETF: the RFC 5170 (2008) defines the codes and their use in FLUTE/ALC, a protocol stack for massively scalable and reliable content delivery services, an active Internet-Draft published last year describes the use of these AL-FEC codes in FECFRAME, a framework for robust real-time streaming applications, and recent Internet-Drafts [91][92] define the GOE (Generalized Object Encoding) extension of LDPC-Staircase codes for UEP (Unequal Erasure Protection) and file bundle protection services.

This success has also been made possible, on the other hand, by our efforts in terms of design and evaluation of two efficient software codecs for LDPC-Staircase codes. One of them is distributed in open-source, as part of our OpenFEC project (http://openfec.org), a unique initiative that aims at promoting open and free AL-FEC solutions. The second one, a highly optimized version with improved decoding speed and reduced memory requirements, is commercialized through an industrial partner, Expway.

Since May 2012, along with the Expway French company, we are proposing the Reed-Solomon + LDPC-Staircase codes for the 3GPP-eMBMS call for technology, as a candidate for next generation AL-FEC codes for multimedia services. We have shown that these codes offer very good erasure recovery capabilities, in line with 3GPP requirements, and extremely high decoding speeds, usually significantly faster than that of the other proposals. The final decision is expected for end of January 2013. In any case we have once again showed that these codes provide very good performance, often ahead of the competitors, and an excellent balance between several technical and non technical criteria.

Finally our activities in the context of the PhD of F. Mattoussi include the design, analysis and improvement of GLDPC-Staircase codes, a "Generalized" extension to LDPC-Staircase codes. We have shown in particular that these codes: (1) offer small rate capabilities, i.e. can produce a large number of repair symbols 'on-the-fly', when needed; (2) feature high erasure recovery capabilities, close to that of ideal codes. Therefore they offer a nice opportunity to extend the field of application of existing LDPC-Staircase codes (IETF RFC 5170), while keeping backward compatibility (i.e. LDPC-Staircase "codewords" can be decoded with a GPLDPC-Staircase codec). More information is available in [56][57][55].

- **Unequal Erasure Protection (UEP) and File bundle protection through the GOE (Generalized Object Encoding) scheme**

  This activity has been initiated with the PostDoc work of Rodrigue IMAD. It focuses on Unequal Erasure Protection capabilities (UEP) (when a subset of an object has more importance than the remaing) and file bundle protection capabilities (e.g. when one want to globally protect a large set of small objects).

  After an in-depth understanding of the well-known PET (Priority Encoding Technique) scheme, and the UOD for RaptorQ (Universal Object Delivery) initiative of Qualcomm, which is a realization of the PET approach, we have designed the GOE FEC Scheme (Generalized Object Encoding) alternative. The idea, simple, is to decouple the FEC protection from the natural object boundaries, and to apply an independant FEC encoding to each "generalized object". The main difficulty is to find an appropriate signaling solution to synchronize the sender and receiver on the exact way FEC encoding is applied. In [91] we show this is feasible, while keeping a backward compatibility with receivers that do not support GOE FEC schemes. Two well known AL-FEC schemes have also been extended to support this new approach, with very minimal modifications, namely Reed-Solomon and LDPC-Staircase codes [92], [91].

  During this work, we compared the GOE and UOD/PET schemes, both from an analytical point of view (we use an N-truncated negative binomial distribution to that purpose) and from an experimental, simulation based, point of view [64]. We have shown that the GOE approach, by the flexibility it offers, its simplicity, its backward compatibility and its good recovery capabilities (under finite of infinite length conditions), outperforms UOD/PET for practical realizations of UEP/file bundle protection systems. See also http://www.ietf.org/proceedings/81/slides/rmt-2.pdf.

- **Application-Level Forward Error Correction Codes (AL-FEC) and their Applications to Robust Streaming Systems**

  AL-FEC codes are known to be useful to protect time-constrained flows. The goal of the IETF FECFRAME working group is to design a generic framework to enable various kinds of AL-FEC schemes to be integrated within RTP/UDP (or similar) data flows. Our contributions in the IETF context are three fold. First of all, we have contributed to the design and standardization of the FECFRAME framework, now published as a Standards Track RFC6363.

  Secondly, we have proposed the use of Reed-Solomon codes (with and without RTP encapsulation of repair packets) and LDPC-Staircase codes within the FECFRAME framework: [85] for Reed-Solomon and [88] for LDPC-Staircase. Both documents are close to being published as RFCs.

Finally, in parallel, we have started an implementation of the FECFRAME framework in order to gain an in-depth understanding of the system. Previous results showed the benefits of LDPC-Staircase codes when dealing with high bit-rate real-time flows.

A second type of activity, in the context of robust streaming systems, consisted in the analysis of the Tetrys approach. Tetrys is a promising technique that features high reliability while being independent from RTT, and performs better than traditional block FEC techniques in a wide range of operational conditions.

- **A new File Delivery Application for Broadcast/Multicast Systems**

  FLUTE [95] has long been the one and only official file delivery application on top of the ALC reliable multicast transport protocol. However FLUTE has several limitations (essentially because the object meta-data are transmitted independently of the objects themselves, in spite of their inter-dependency), features an intrinsic complexity, and is only available for ALC.

  Therefore, we started the design of FCAST, a simple, lightweight file transfer application, that works both on top of both ALC and NORM [82]. This work is carried out as part of the IETF RMT Working Group, in collaboration with B. Adamson (NRL). This document has passed WG Last Call and is currently considered by IESG.

- **Security of the Broadcast/Multicast Systems**

  Sooner or later, broadcasting systems will require security services. This is all the more true as heterogeneous broadcasting technologies are used, some of them being by nature open, such as WiFi networks. Therefore, one of the key security services is the authentication of the packet origin and the packet integrity check. To that purpose, we have specified the use of simple authentication and integrity schemes (i.e., group MAC and digital signatures) in the context of the ALC and NORM protocols and the standard is now published as IETF RFC 6584 [98].

- **High Performance Security Gateways for High Assurance Environments**

  This work focuses on very high performance security gateways, compatible with 10Gbps or higher IPsec tunneling throughput, while offering a high assurance thanks in particular to a clear red/black flow separation. In this context we have studied last year the feasibility of high-bandwidth, secure communications on generic machines equipped with the latest CPUs and General-Purpose Graphical Processing Units (GPGPU).

  The work carried out in 2011-2012 consisted in setting up and evaluating the high performance platform. This platform heavily relies on the Click modular TCP/IP protocol stack implementation, which turned out to be a key enabler both in terms of specialization of the stack and parallel processing. Our activities also consisted in analyzing the PMTU discovery aspect since it is a critical factor in achieving high bandwidths. To that goal we have designed a new approach for qualifying ICMP blackholes in the Internet, since PMTUD heavily relies on ICMP [51].

6.2. Network Security and Privacy

Participants: Claude Castelluccia, Gergely Acs, Mathieu Cunche, Daniele Perito, Lukasz Olejnik, Mohamed Ali Kaafar, Abdelberi Chaabane, Cédric Lauradoux, Minh-Dung Tran.

- *Private Big Data Publication* Public datasets are used in a variety of applications spanning from genome and web usage analysis to location-based and recommendation systems. Publishing such datasets is important since they can help us analyzing and understanding interesting patterns. For example, mobility trajectories have become widely collected in recent years and have opened the possibility to improve our understanding of large-scale social networks by investigating how people exchange information, interact, and develop social interactions. With billion of handsets in use worldwide, the quantity of mobility data is gigantic. When aggregated, they can help understand complex processes, such as the spread of viruses, and build better transportation systems, prevent traffic congestion. While the benefits provided by these datasets are indisputable, they unfortunately pose a considerable threat to individual privacy. In fact, mobility trajectories might be used by a malicious attacker to discover potential sensitive information about a user, such as his habits, religion or relationships. Because privacy is so important to people, companies and researchers are reluctant to publish datasets by fear of being held responsible for potential privacy breaches. As a result, only very few of them are actually released and available. This limits our ability to analyze such data to derive information that could benefit the general public. Here follows some recent results of our activities in this domain.

  Privacy-Preserving Sequential Data Publication [41]: Sequential data is being increasingly used in a variety of applications, spanning from genome and web usage analysis to location-based recommendation systems. Publishing sequential data is of vital importance to the advancement of these applications since they can enable researchers to analyze and understand interesting sequential patterns. However, as shown by the re-identification attacks on the AOL and Netflix datasets, releasing sequential data may pose considerable threats to individual privacy. Recent research has indicated the failure of existing sanitization techniques to provide claimed privacy guarantees. It is therefore urgent to respond to this failure by developing new schemes with provable privacy guarantees. Differential privacy is one of the only models that can be used to provide such guarantees. Due to the inherent sequentiality and high-dimensionality, it is challenging to apply differential privacy to sequential data. In this work, we address this challenge by employing a variable-length n-gram model, which extracts the essential information of a sequential database in terms of a set of variable-length n-grams. Our approach makes use of a carefully designed exploration tree structure and a set of novel techniques based on the Markov assumption in order to lower the magnitude of added noise. The published n-grams are useful for many purposes. Furthermore, we develop a solution for generating a synthetic database, which enables a wider spectrum of data analysis tasks. Extensive experiments on real-life datasets demonstrate that our approach substantially outperforms the state-of-the-art techniques.

  Private Histogram Publishing [33]:

  Differential privacy can be used to release different types of data, and, in particular, histograms, which provide useful summaries of a dataset. Several differentially private histogram releasing schemes have been proposed recently. However, most of them directly add noise to the histogram counts, resulting in undesirable accuracy. In this work, we propose two sanitization techniques that exploit the inherent redundancy of real-life datasets in order to boost the accuracy of histograms. They lossily compress the data and sanitize the compressed data. Our first scheme is an optimization of the Fourier Perturbation Algorithm (FPA) presented in [13]. It improves the accuracy of the initial FPA by a factor of 10. The other scheme relies on clustering and exploits the redundancy between bins. Our extensive experimental evaluation over various real-life and synthetic datasets demonstrates that our techniques preserve very accurate distributions and considerably improve the accuracy of range queries over attributed histograms.

- *Privacy Issues on the Internet* Internet users are being increasingly tracked and profiled. Companies utilize profiling to provide customized, i.e. personalized services to their customers, and hence increase revenues.

  Privacy issues of Targeted Advertising [37]: Behavioral advertising takes advantage from profiles of users' interests, characteristics (such as gender, age and ethnicity) and purchasing activities. For example, advertising or publishing companies use behavioral targeting to display advertisements that closely reflect users' interests (e.g. 'sports enthusiasts'). Typically, these interests are inferred from users' web browsing activities, which in turn allows building of users' profiles. It can be argued that customization resulting from profiling is also beneficial to users who receive useful information and relevant online ads in line with their interests. However, behavioral targeting is often perceived as a threat to privacy mainly because it heavily relies on users' personal information, collected by only a few companies. In this work, we show that behavioral advertising poses an additional privacy threat because targeted ads expose users' private data to any entity that has access to a small portion of these ads. More specifically, we show that an adversary who has access to a user's targeted ads can retrieve a large part of his interest profile. This constitutes a privacy breach because interest profiles often contain private and sensitive information.

  On the Uniqueness of Web Browsing History Patterns [60]: We present the results of the first large-scale study of the uniqueness of Web browsing histories, gathered from a total of $368,284$ Internet users who visited a history detection demonstration website. Our results show that for a majority of users (69%), the browsing history is unique and that users for whom we could detect at least 4 visited websites were uniquely identified by their histories in 97% of cases. We observe a high rate of stability in browser history fingerprints: for repeat visitors, 80% of fingerprints are identical over time, and differing ones were strongly correlated with original history contents, indicating static browsing preferences. We report a striking result that it is enough to test for a small number of pages in order to both enumerate users' interests and perform an efficient and unique behavioral fingerprint; we show that testing 50 web pages is enough to fingerprint 42% of users in our database, increasing to 70% with 500 web pages. Finally, we show that indirect history data, such as information about *categories* of visited websites can also be effective in fingerprinting users, and that similar fingerprinting can be performed by common script providers such as Google or Facebook.
- **Adaptive Password-Strength Meters from Markov Models [38]**

  Passwords are a traditional and widespread method of authentication, both on the Internet and off-line. Passwords are portable, easy to understand for laypersons, and easy to implement for the operator. Thus, password-based authentication is likely to stay for the foreseeable future.

  To ensure an acceptable level of security of user-chosen passwords, sites often use mechanisms to test the strength of a password (often called *pro-active password checkers*) and then reject weak passwords. Hopefully this ensures that passwords are reasonably strong on average and makes guessing passwords infeasible or at least too expensive for the adversary. Commonly used password checkers rely on rules such as requiring a number and a special character to be used. However, as we will show and also has been observed in previous work, the accuracy of such password checkers is low, which means that often insecure passwords are accepted and secure passwords are rejected. This adversely affects both security and usability.

  In this work, we propose to use password strength meters based on Markov-models, which estimate the true strength of a password more accurately than rule-based strength meters. Roughly speaking, the Markov-model estimates the strength of a password by estimating the probability of the n-grams that compose said password. Best results can be obtained when the Markov-models are trained on the actual password database. We show, in this work, how to do so without sacrificing the security of the password database, even when the n-gram database is leaked.

  We show how to build secure adaptive password strength meters, where security should hold even when the n-gram database leaks. This is similar to traditional password databases, where one tries to minimize the effects of a database breach by hashing and salting the stored passwords. This is not a trivial task. One potential problem is that, particularly strong passwords, can be leaked entirely by an n-gram database (without noise added).

- **Fast Zero-Knowledge Authentication [47]** We explore new area/throughput trade-offs for the Girault, Poupard and Stern authentication protocol (GPS). This authentication protocol was selected in the NESSIE competition and is even part of the standard ISO/IEC 9798. The originality of our work comes from the fact that we exploit a fixed key to increase the throughput. It leads us to implement GPS using the Chapman constant multiplier. This parallel implementation is 40 times faster but 10 times bigger than the reference serial one. We propose to serialize this multiplier to reduce its area at the cost of lower throughput. Our hybrid Chapman's multiplier is 8 times faster but only twice bigger than the reference. Results presented here allow designers to adapt the performance of GPS authentication to their hardware resources. The complete GPS prover side is also integrated in the network stack of the PowWow sensor which contains an Actel IGLOO AGL250 FPGA as a proof of concept.
- **Energy Efficient Authentication Strategies for Network Coding [26]**

  Recent advances in information theory and networking, e.g. aggregation, network coding or rateless codes, have significantly modified data dissemination in wireless networks. These new paradigms create new threats for security such as pollution attacks and denial of services (DoS). These attacks exploit the difficulty to authenticate data in such contexts. The particular case of xor network coding is considered herein. We investigate different strategies based on message authentication codes algorithms (MACs) to thwart these attacks. Yet, classical MAC designs are not compatible with the linear combination of network coding. Fortunately, MACs based on universal hash functions (UHFs) match nicely the needs of network coding: some of these functions are linear $h(x_1 \oplus x_2) = h(x_1) \oplus h(x_2)$. To demonstrate their efficiency, we consider the case of wireless sensor networks (WSNs). Although these functions can drastically reduce the energy consumption of authentication (up to 68% gain over the classical designs is observed), they increase the threat of DoS. Indeed, an adversary can disrupt all communications by polluting few messages. To overcome this problem, a group testing algorithm is introduced for authentication resulting in a complexity linear in the number of attacks. The energy consumption is analyzed for cross-point and butterfly network topologies with respect to the possible attack scenarios. The results highlight the trade-offs between energy efficiency, authentication and the effective throughput for the different MAC modes.
- **Towards Stronger Jamming Model: Application to TH-UWB Radio [35]**

  With the great expansion of wireless communications, jamming becomes a real threat.We propose a new model to evaluate the robustness of a communication system to jamming. The model results in more scenarios to be considered ranging from the favorable case to the worst case. The model is applied to a TH-UWB radio. The performance of such a radio in presence of the different jamming scenarios is analyzed. We introduce a mitigation solution based on stream cipher that restricts the jamming problem of the TH-UWB communication to the more favorable case while preserving confidentiality.
- **Privacy risks quantification in Online social networks**

  In this project, we analyze the different capabilities of online social networks and aim to quantify the privacy risks users are undertaking in this context. Online Social Networks (OSNs) are a rich source of information about individuals. It may be difficult to justify the claim that the existence of public profiles breaches the privacy of their owners, as they are the ones who entered the data and made them publicly available in the first place. However, aggregation of multiple OSN public profiles is debatably a source of privacy loss, as profile owners may have expected each profile's information to stay within the boundaries of the OSN service in which it was created. First we present an empirical study of personal information revealed in public profiles of people who use multiple Online Social Networks (OSNs). This study aims to examine how users reveal their personal information across multiple OSNs. We consider the number of publicly available attributes in public profiles, based on various demographics and show a correlation between the amount of information revealed in OSN profiles and specific occupations and the use of pseudonyms. Then, we measure the complementarity of information across OSNs and contrast it with our observations about users who share a larger amount of information. We also measure the consistency of information revelation patterns across OSNs, finding that users have preferred patterns when revealing information across OSNs. To evaluate the quality of aggregated profiles we introduce a consistency measure for attribute values, and show that aggregation also improves information granularity. Finally, we demonstrate how the availability of multiple OSN profiles can be exploited to improve the success of obtaining users' detailed contact information, by cross-linking with publicly available data sources such as online phone directories. This work has been published in ACM SIGCOMM WOSN [42].

In a second study, we examine the user tracking capabilities of the three major global Online Social Networks (OSNs). We study the mechanisms which enable these services to persistently and accurately follow users web activity, and evaluate to which extent this phenomena is spread across the web. Through a study of the top 10K websites, our findings indicate that OSN tracking is diffused among almost all website categories, independently from the content and from the audience. We also evaluate the tracking capabilities in practice and demonstrate by analyzing a real traffic traces that OSNs can reconstruct a significant portion of users web profile and browsing history. We finally provide insights into the relation between the browsing history characteristics and the OSN tracking potential, highlighting the high risk properties. This work has also been published in ACM SIGCOMM WOSN [40].

In a third study, we also analyzed the inference capabilities of third parties from seemingly harmless and unconsciously publicly shared data. Interests (or "likes") of users is one of the highly-available on-line information on the web. In this study, we show how these seemingly harmless interests (e.g., music interests) can leak privacy sensitive information about users. In particular, we infer their undisclosed (private) attributes using the public attributes of other users sharing similar interests. In order to compare user-defined interest names, we extract their semantics using an ontologized version of Wikipedia and measure their similarity by applying a statistical learning method. Besides self-declared interests in music, our technique does not rely on any further information about users such as friend relationships or group belongings. Our experiments, based on more than 104K public profiles collected from Facebook and more than 2000 private profiles provided by volunteers, show that our inference technique efficiently predicts attributes that are very often hidden by users. This is the first time that user interests are used for profiling, and more generally, semantics-driven inference of private data is addressed. Our work received many media attention and was published in the prestigious NDSS symposium [39].

- **On the Privacy threats of hidden information in Wireless communication**

  Wi-Fi protocol has the potential to leak personal information. Wi- Fi capable devices commonly use active discovery mode to find the available Wi-Fi access points (APs). This mechanism includes broadcast of the AP names to which the mobile device has previously been connected to, in plain text, which may be easily observed and captured by any Wi-Fi device monitoring the control traffic. The combination of the AP names belonging to any mobile device can be considered as a Wi-Fi fingerprint, which can be used to identify the mobile device user. Our research investigates how it is possible to exploit these fingerprints to identify links between users i.e. owners of the mobile devices broadcasting such links. In this project, we have used an approach based on the similarity between the Wi-Fi fingerprints, which is equated to the likelihood of the corresponding users being linked. When computing the similarity between two Wi-Fi fingerprints, two dimensions need to be considered : (i) The number of network names in common. Indeed, sharing a network is an indication of the existence of a link, e.g. friends and family that share multiple Wi-Fi networks. (ii) The rarity of the network names in common. Some network names are very common and sharing them does not imply a link between the users. This is the case for public network names such as McDonalds Free Wi-Fi, or default network names such as NETGEAR and Linksys. On the other hand, uncommon network names such as Griffin Family Network or Orange-3EF50 are likely to indicate a strong link between the users of these networks. Utilising a carefully designed similarity metric, we have been able to infer the existence of social links with a high confidence: 80% of the links were detected with an error rate of 7%. We show that through real-life experiments that owners of smartphones are particularly exposed to this threat, as indeed these devices are carried on persons throughout the day, connecting to multiple Wi-Fi networks and also broadcasting their connection history. There are a number of industry and research initiatives aiming to address Wi-Fi related privacy issues. The deployment of new technology i.e. privacy preserving discovery services, would necessitate software modifications in currently deployed APs and devices. The obvious solution to disable active discovery mode, comes at the expense of performance and usability, i.e. with an extended time duration for the Wi-Fi capable device to find and connect to an available AP. As a possible first step, users should be encouraged to remove the obsolete connection history entries, which may lower the similarity metric and thus reduce the ease of linkage. Our papers illustrating this study have been presented in the WoWMoM'12 conference [45] and in the IEEE MILCOM conference [43].

- **Information leakage in Ads networks**

  In targeted (or behavioral) advertising, users' behaviors are tracked over time in order to customize served ads to their interests. This creates serious privacy concerns since for the purpose of profiling, private information is collected and centralized by a limited number of companies. Despite claims that this information is secure, there is a potential for this information to be leaked through the customized services these companies are offering. In this study, we show that targeted ads expose users' private data not only to ad providers but also to any entity that has access to users' ads. We propose a methodology to filter targeted ads and infer users' interests from them. We show that an adversary that has access to only a small number of websites containing Google ads can infer users' interests with an accuracy of more than 79% (Precision) and reconstruct as much as 58% of a Google Ads profile in general (Recall). This study is the first work that identifies and quantifies information leakage through ads served in targeted advertising. We published a paper illustrating these results in the prestigious Privacy Enhancing Technologies Symposium PETS 2012 [37].

- **Privacy in P2P file sharing systems**

  In this study, we aim at characterizing anonymous file sharing systems from a privacy perspective. We concentrate on a recently deployed privacy-preserving file sharing system: OneSwarm. Our characterisation is based on measurement of several aspects of the OneSwarm system such as the nature of the shared and searched content and the geolocation and number of users. Our findings indicate that, as opposed to common belief, there is no significant difference in downloaded content between this system and the classical BitTorrent ecosystem. We also found that a majority of users appear to be located in countries where anti-piracy laws have been recently adopted and enforced (France, Sweden and U.S). Finally, we evaluate the level of privacy provided by OneSwarm, and show that, although the system has strong overall privacy, a collusion attack could potentially identify content providers. This work has been published in [46].

- **Privacy leakage on mobile devices: the Mobilitics Inria-CNIL project**

  This joint Inria-CNIL (the French data protection agency) project aims at assessing the privacy risks associated to the use of smartphones and tablets, in particular because of personal information leakage to remote third parties. Both applications and the base OS services are considered as potential source of information leakage. More precisely, the goals are to define a platform and a methodology to identify, measure, and see the evolution over the time of privacy risks.

  If similar risks exist with a PC, the situation is more worrying with mobile terminals. The reasons are:

  – the intrusive feature of these terminals that their owner continuously keep with them;
  – the amount of personnal information available on these terminals (mobile terminals aggregate personnal information but also create them, for instance with geolocalisation information);

- the facility with which the owner can personnalize its terminal with new applications;
- the financial incentives that lead companies to collect and use personnal information;
- the fact that the terminal user has no tool (e.g. a "privacy" firewall) to control precisely what information is exchanged with whom. The permissions provided by Android is too coarse grained to be useful, and the new privacy dashboard of IOS 6 does not enable the user to have an idea of how personnal information is used by an authorized application (a one time access to a personnal information and local processing within the application can be acceptable, whereas the periodic transmission of this information to remote servers is not);

The final goals of the Mobilitics project are both to study the situation and trend, but also to make mobile terminal users aware of the situation, and to provide tools that may help them to better control the personnal information flow of their terminal.

6.3. Formal and legal issues of privacy

Participants: Thibaud Antignac, Denis Butin, Daniel Le Métayer.

- **Verification of privacy properties** The increasing official use of security protocols for electronic voting deepens the need for their trustworthiness, hence for their formal verification. The impossibility of linking a voter to her vote, often called voter privacy or ballot secrecy, is the core property of many such protocols. Most existing work relies on equivalence statements in cryptographic extensions of process calculi. We have proposed the first theorem-proving based verification of voter privacy which overcomes some of the limitations inherent to process calculi-based analysis [36]. Unlinkability between two pieces of information is specified as an extension to the Inductive Method for security protocol verification in Isabelle/HOL. New message operators for association extraction and synthesis are defined. Proving voter privacy demanded substantial effort and provided novel insights into both electronic voting protocols themselves and the analysed security goals. The central proof elements have been shown to be reusable for different protocols with minimal interaction.
- **Privacy by design** The privacy by design approach is often praised by lawyers as well as computer scientists as an essential step towards a better privacy protection. The general philosophy of privacy by design is that privacy should not be treated as an afterthought but rather as a first-class requirement during the design of a system. The approach has been applied in different areas such as smart metering, electronic traffic pricing, ubiquitous computing or location based services. More generally, it is possible to identify a number of core principles that are widely accepted and can form a basis for privacy by design. For example, the Organization for Economic Co-operation and Development (OECD) has put forward principles such as the consent, limitation of use, data quality, security and accountability. One must admit however that the take-up of privacy by design in the industry is still rather limited. This situation is partly due to legal and economic reasons: as long as the law does not impose binding commitments, ICT providers and data collectors do not have sufficient incentives to invest into privacy by design. The situation on the legal side might change in Europe though because the regulation proposed by the European Commission in January 2012 (to replace the European Directive 95/46/EC) includes binding commitments on privacy by design.

  But the reasons for the lack of adoption of privacy by design are not only legal and economic: even though computer scientists have devised a wide range of privacy enhancing tools, no general methodology is available to integrate them in a consistent way to meet a set of privacy requirements. The next challenge in this area is thus to go beyond individual cases and to establish sound foundations and methodologies for privacy by design. As a first step in this direction, we have focused on the data minimization principle which stipulates that the collection should be limited to the pieces of data strictly necessary for the purpose, and we have proposed a framework to reason about the choices of architecture and their impact in terms of privacy [53]. The first strategic choices are the allocation of the computation tasks to the nodes of the architecture and the types of communications between the nodes. For example, data can be encrypted or hashed, either to protect their confidentiality or to provide guarantees with respect to their correctness or origin. The main benefit of a centralized architecture for the “central” actor is that he can trust the result because he keeps full control over its computation. However, the loss of control by a single actor in decentralized architectures can be offset by extra requirements ensuring that errors (or frauds) can be detected *a posteriori*. In order to help the designer grasp the combination of possible options, our framework provides means to express the parameters to be taken into account (the service to be performed, the actors involved, their respective requirements, etc.) and an inference system to derive properties such as the possibility for an actor to detect potential errors (or frauds) in the computation of a variable. This inference system can be used in the design phase to check if an architecture meets the requirements of the parties or to point out conflicting requirements.

- **Privacy and discrimination**

  Actually, the interactions between personal data protection, privacy and protection against discriminations are increasingly numerous and complex. For example, there is no doubt that misuses of personal data can adversely affect privacy and self-development (for example, resulting in the unwanted disclosure of personal data to third parties, in identity theft, or harassment through email or phone calls), or lead to a loss of choices or opportunities (for example, enabling a recruiter to obtain information over the internet about political opinions or religious beliefs of a candidate and to use this information against him). It could even be suggested that privacy breaches and discriminations based on data processing are probably the two most frequent and the most serious types of consequences of personal data breaches. We have studied these interactions from a multidisciplinary (legal and technical) perspective and argued that an extended application of the application of non-discrimination regulations could help strengthening data protection [52]. We have analysed and compared personal data protection, privacy and protection against discriminations considering both the types of data concerned and the *modus operandi* (*a priori* versus *a posteriori* controls, actors in charge of the control, etc.). From this comparison, we have drawn some conclusions with respect to their relative effectiveness and argued that *a posteriori* controls on the use of personal data should be strengthened and the victims of data misuse should get compensations which are significant enough to represent a deterrence for data controllers. We have also advocated the establishment of stronger connections between anti-discrimination and data protection laws, in particular to ensure that any data processing leading to unfair differences of treatments between individuals is prohibited and can be effectively punished [29].

6.4. Network measurement, modeling and understanding

Participants: Chadi Barakat, Arnaud Legout, Ashwin Rao, Walid Dabbous, Tessema Mindaye, Mohamed Ali Kaafar, Dong Wang, Vincent Roca, Ludovic Jacquin, Byungchul Park.

The main objective of our work in this domain is a better monitoring of the Internet and a better understanding of its traffic. We work on new measurement techniques that scale with the fast increase in Internet traffic and growth of its size. We propose solutions for a fast and accurate identification of Internet traffic based on packet size statistics and host profiles. Within the ANR CMON project, we work on monitoring the quality of the Internet access by end-to-end probes, and on the detection and troubleshooting of network problems by collaboration among end users.

Next, is a sketch of our main contributions in this area.

- **Checking Traffic Differentiation at the Internet Access**

  In the last few years, ISPs have been reported to discriminate against specific user traffic, especially if generated by bandwidth-hungry applications. The so-called network neutrality, advocating that an ISP should treat all incoming packets equally, has been a hot topic ever since. We propose Chkdiff, a novel method to detect network neutrality violations that takes a radically different approach from existing work: it aims at both application and differentiation technique agnosticism. We achieve this in three steps. Firstly, we perform measurements with the user’s real traffic instead of using specific application traces. Secondly, we do not assume that discrimination takes place on any particular packet field, which requires us to preserve the integrity of all the traffic we intend to test. Thirdly, we detect differentiation by comparing the performance of a traffic flow against that of all other traffic flows from the same user, considered as a whole.

Chkdiff is based on the following key ideas:
Idea 1: **Use real user traffic**. We want to test the existence of traffic discrimination for the exact set of applications run by the end user. Hence, we only consider user-generated traffic.
Idea 2: **Leave user traffic unchanged, or almost**. All methods performing active measurements send probes made of real application packets and of packets that are similar, but slightly modified, so that they do not get discriminated along their path. This is quite an assumption, as we do not know exactly what ISPs do behind the scenes. In the extreme case, ISPs could even white-list traffic generated by differentiation detecting tools. It is therefore crucial to preserve as much of the original packets as possible, as well as their original per-flow order. We will see that the modifications introduced by our tool affect only the ordering of packets, their TTL value or their IP identification field.
Idea 3: **Baseline is the entire traffic performance**. Since we do not want to make any hypothesis in advance on what kind of mechanisms - if any - are deployed, we claim that the performance of each single non-differentiated flow should present the same behaviour as that of the rest of our traffic as a whole. Differentiated flows, on the other hand, should stand out when compared to all other flows grouped together, where a large fraction of non-differentiated flows should mitigate the impact of differentiated ones.

Chkdiff is currently the subject of a collaboration with I3S around the PhD thesis of Riccardo Ravaioli (funded by the Labex UCN@Sophia). A first description of the tool is presented in [63].

- **Lightweight Enhanced Monitoring for High-Speed Networks**

  Within the collaboration with Politecnico di Bari, we worked on LEMON, a lightweight enhanced monitoring algorithm based on packet sampling. This solution targets a pre-assigned accuracy on bitrate estimates, for each monitored flow at a router interface. To this end, LEMON takes into account some basic properties of the flows, which can be easily inferred from a sampled stream, and exploits them to dynamically adapt the monitoring time-window on a per-flow basis. Its effectiveness is tested using real packet traces. Experimental results show that LEMON is able to finely tune, in real-time, the monitoring window associated to each flow and its communication overhead can be kept low enough by choosing an appropriate aggregation policy in message exporting. Moreover, compared to a classic fixed-scale monitoring approach, it is able to better satisfy the accuracy requirements of bitrate estimates. Finally, LEMON incurs a low processing overhead, which can be easily sustained by currently deployed routers, such as a CISCO 12000 device. This work is currently under submission.

- **The Complete Picture of the Twitter Social Graph**

  In this work [49], we collected the entire Twitter social graph that consists of 537 million Twitter accounts connected by 23.95 billion links, and performed a preliminary analysis of the collected data. In order to collect the social graph, we implemented a distributed crawler on the PlanetLab infrastructure that collected all information in 4 months. Our preliminary analysis already revealed some interesting properties. Whereas there are 537 million Twitter accounts, only 268 million already sent at least one tweet and no more than 54 million have been recently active. In addition, 40% of the accounts are not followed by anybody and 25% do not follow anybody. Finally, we found that the Twitter policies, but also social conventions (like the followback convention) have a huge impact on the structure of the Twitter social graph.

- **Meddle: Middleboxes for Increased Transparency and Control of Mobile Traffic**

  Mobile networks are the most popular, fastest growing and least understood systems in today's Internet ecosystem. Despite a large collection of privacy, policy and performance issues in mobile networks users and researchers are faced with few options to characterize and address them. In this work [62] we designed Meddle, a framework aimed at enhancing transparency in mobile networks and providing a platform that enables users (and researchers) control mobile traffic. In the mobile environment, users are forced to interact with a single operating system tied to their device, generally run closedsource apps that routinely violate user privacy, and subscribe to network providers that can (and do) transparently modify, block or otherwise interfere with network traffic. Researchers face a similar set of challenges for characterizing and experimenting with mobile systems. To characterize mobile traffic and design new protocols and services that are better tailored to the mobile environment, we would like a framework that allows us to intercept and potentially modify traffic generated by mobile devices as they move with users, regardless of the device, OS, wireless technology, or carrier. However, implementing this functionality is difficult on mobile devices because it requires warrantyvoiding techniques such as jail breaking to access and manipulate traffic at the network layer. Even when using such an approach, carriers may manipulate traffic once it leaves the mobile device, thus rendering some research impractical. Furthermore, researchers generally have no ability to deploy solutions and services such as prefetching and security filters, that should be implemented in the network. In this work, we designed Meddle, a framework that combines virtual private networks (VPNs) with middleboxes to provide an experimental platform that aligns the interests of users and researchers.

- **Mobile users' behavior modeling in Video on Demand systems and its implication on user privacy and caching strategies**

  In this project, we examine mobile users' behavior and their corresponding video viewing patterns from logs extracted from the servers of a large scale VoD system. We focus on the analysis of the main discrepancies that might exist when users access the VoD system catalog from WiFi or 3G connections. We also study factors that might impact mobile users' interests and video popularity. The users' behavior exhibits strong daily and weekly patterns, with mobile users' interests being surprisingly spread across almost all categories and video lengths, independently of the connection type. However, by examining the activity of users individually, we observed a concentration of interests and peculiar access patterns, which allows to classify the users and thus better predict their behavior. We also find the skewed video popularity distribution and demonstrate that the popularity of a video can be predicted using its very early popularity level. We then analyzed the sources of video viewing and found that even if search engines are the dominant sources for a majority of videos, they represent less than 10% (resp. 20%) of the sources for the highly popular videos in 3G (resp. WiFi) network. We also report that both the type of connection and the type of mobile device used have an impact on the viewing time and the source of viewing. Using our findings, we provide insights and recommendations that can be used to design intelligent mobile VoD systems and help in improving personalized services on these platforms. This work has been published in IMC 2012 [54].

- **Explicative models for Information Spreading on the web from a user profiling perspective**

Microblog services offer a unique approach to online information sharing allowing microblog users to forward messages to others. We study the process of information diffusion in a microblog service developing Galton-Watson with Killing (GWK) model, which has many implications ranging from privacy protection to experiments validation and benchmarking. We describe an information propagation as a discrete GWK process based on Galton-Watson model which models the evolution of family names. Our model explains the interaction between the topology of the social graph and the intrinsic interest of the message. We validate our models on dataset collected from Sina Weibo and Twitter microblogs. Sina Weibo is a Chinese microblog web service which reached over 100 million users as for January 2011. Our Sina Weibo dataset contains over 261 thousand tweets which have retweets and 2 million retweets from 500 thousand users. Twitter dataset contains over 1.1 million tweets which have retweets and 3.3 million retweets from 4.3 million users. The results of the validation show that our proposed GWK model fits the information diffusion of microblog service very well in terms of the number of message receivers. We show that our model can be used in generating tweets load and also analyze the relationships between parameters of our model and popularity of the diffused information. Our work is the first to give a systemic and comprehensive analysis for the information diffusion on microblog services, to be used in tweets-like load generators while still guaranteeing popularity distribution characteristics. Our paper illustrating this study will be presented in IEEE Infocom 2013 [69].

- **Tracking ICMP black holes at an Internet Scale**

  ICMP is a key protocol to exchange control and error messages over the Internet. An appropriate ICMP's processing throughout a path is therefore a key requirement both for troubleshooting operations (e.g. debugging routing problems) and for several functionalities (e.g. Path Maximum Transmission Unit Discovery, PMTUD). Unfortunately it is common to see ICMP malfunctions, thereby causing various levels of problems. In our study, we first introduce a taxonomy of the way routers process ICMP, which is of great help to understand for instance certain traceroute outputs. Secondly we introduce IBTrack, a tool that any user can use to automatically characterize ICMP issues within the Internet, without requiring any additional in-network assistance (e.g. there is no vantage point). Finally we validate our IBTrack tool with large scale experiments and we take advantage of this opportunity to provide some statistics on how ICMP is managed by Internet routers. This work has been presented in IEEE Globecom [51].

6.5. Experimental Environment for Future Internet Architecture

Participants: Walid Dabbous, Thierry Parmentelat, Fréderic Urbani, Daniel Camara, Alina Quereilhac, Shafqat Ur-Rehman, Mohamed Larabi, Thierry Turletti, Julien Tribino.

- **SFA Federation of experimental testbeds**

  We are now involved in the NOVI (E.U. STREP) project, the F-Lab (French A.N.R.) project, the FED4FIRE (E.U. IP) project and have the lead of the "Control Plane Extensions" WorkPackage of OpenLab (E.U. IP) project. Within these frameworks, as part of the co-development agreement between the Planète team and Princeton University, we have made a great deal of contributions into one of the most visible and renown implementations of the Testbed-Federation architecture known as SFA for Slice-based Federation Architecture. As a sequel of former activities we also keep a low-noise maintenance activity of the PlanetLab software, which has been running in particular on the PlanetLab global testbed since 2004, with an ad-hoc federated model in place between PlanetLab Central (hosted by Princeton University) and PlanetLab Europe (hosted at Inria) since 2007.

  During 2012, we have focused on the maturation of the SFA specifications and the SfaWrap codebase, with several objectives in mind. Firstly, we have contributed within the GENI (N.S.F.) project to the specifications of the Version 3 of the AM-API (Aggregate Manager API), which defines the primitives that a testbed management infrastructure has to provide in order to be SFA-compliant.

Secondly, knowing that our former SFA implementation was targeting PlanetLab testbeds only, we needed on the one hand, to make generic this SFA implementation, by completely redesign and refactor its codebase, and on the other hand, we needed to support all the resources allocation strategies supported by the testbeds, namely the allocation of both 'shared' and 'exclusive' resources. As a result of this redesign and development effort, out new SFA implementation is now disseminated and started to be known, under the name of SfaWrap, and we believe that it can be used as a production-grade alternative to quickly add SFA compatibility on top of many heterogenous testbed management frameworks.

Finally, in order to allow the community of networking researchers to execute cross-testbed experiments, involving heterogeneous resources, Planète team has been instrumental in federating a set of well-known testbeds through the SfaWrap, namely PlanetLab Europe, Senslab - developed in other Inria Project-teams -, FEDERICA, the outcome of another E.U.-funded project and more recently NITOS, an OMF-enabled wireless testbed. See [96] and [97] for more details.

- **Content Centric Networks Simulation**

  We worked this year on the extension of the DCE framework for ns-3 in order to run CCN implementation under the ns-3 simulator. DCE stands for Direct Code Execution, its goal is to execute unmodified C/C++ binaries under ns-3 network simulator. With this tool researchers and developpers can use the same code to do simulation and real experiments. DCE operation principle is to catch the standard systems calls done by the real application in the experiment and to emulate them within the ns-3 virtual network topology. Concerning CCN we use the PARC implementation named CCNx which is a well working open source software reference implementation of Content Centric Network protocol. As promised by DCE this integration of CCNx requires no modification of its code, it requires 'only' working on adding the system calls used by CCN that are not already supported by DCE. The advantage of this approach is that the integration work of CCN advanced DCE and will be useful in others completly different experiments. Another great advantage is that every evolution of the CCNx implementation is very easy to integrate, all what is needed is to compile the new source code. The next steps will be naturally to use DCE/ns-3 to evaluation CCN protocols in specific scenarii, to improve the coverage of systems calls supported by DCE, and to improve the DCE scheduler to be more realistic and to take into account CPU time spent in router queues. This work is done in the context of the ANR CONNECT project and is currently under submission.

- **ns-3 Module store**

  Bake is an integration tool which is used by software developers to automate the reproducible build of a number of projects which depend on each other and which might be developed, and hosted by unrelated parties. This software is being developed with the participation of the Planète group and is intended to be the automatic building tool adopted by the ns-3 project.

  The client version of Bake is already working and the Planète group had a significant participation in its development. The contributions were in the context the addition of new functionalities, bug fixing and in the development of the regression tests. We are now starting the development of the ns-3 modules repository, which is a web portal to store the meta-information of the available modules. In the present state we have already designed and implemented the portal data basis and the main interface. It is already possible to register new modules and browse among the already registered ones.

The web portal has to be finished, notably the part that will create the xml file that will be used to feed the bake's client. We also need to add new functionalities to the client part, to enable incremental build over partially deployed environments. As it is today, bake does not enable the user to add just one new module to an already deployed version of the ns-3 simulator. This work is done in the context of the ADT MobSim in collaboration with Hipercom and Swing Inria project-teams. For more details see the Bake web page http://planete.inria.fr/software/bake/index.html

- **The ns-3 consortium**

  We have founded last year a consortium between Inria and University of Washington. The goals of this consortium are to (1) provide a point of contact between industrial members and the ns-3 project, to enable them to provide suggestions and feedback about technical aspects, (2) guarantee maintenance of ns-3's core, organize public events in relation to ns-3, such as users' day and workshops and (3) provide a public face that is not directly a part of Inria or NSF by managing the http://www.nsnam.org web site.

- **Automated Deployment and Customization of Routing Overlays Across Heterogeneous Experimentation Platforms**

  During the last decades, many institutions and companies around the world have invested great effort into building new network experimentation platforms. These platforms range from simulators, to emulators and live testbeds, and provide very heterogeneous ways to access resources and to run experiments.

  Currently, a growing concern among platform owners is how to encourage researchers from different platform communities to take advantage of the resources they offer. However, one important aspect that needs to be overcome in order to appeal researchers to use as many experimentation platforms as necessary to best validate their results, is to decrease the inherent complexity to run experiments in different platforms. Even more so, to decrease the complexity of mixing resources from different platforms on a same experiment, to achieve the combination of resources best suited to the experiment needs.

  To address this concern, we developed the Network Experiment Programming Interface (NEPI) whose goal is to make easier the use of different experimentation platforms, and switch among them easily. The development of NEPI started in 2009 with the implementation of the core API, an address allocator, a routing table configurator, but also a prototype ns-3 backend driven by a simple graphical user interface based on QT. On 2010 we validated and evolved the core API with the addition of a new backend based on linux network namespace containers and stabilized the existing ns-3 backend.

  During 2011, we enhanced the design of NEPI and provided experiment validation, distributed experiment control, and failure recovery functionalities. In particular, we enforced separation between experiment design and execution stages, with off-line experiment validation. We also introduced a hierarchical distributed monitoring scheme to control experiment execution. We implemented a stateless message-based communication scheme, and added failure recovery mechanisms to improve robustness. Also on 2011, we started work on a prototype PlanetLab backend.

  Last year, we extended NEPI to provide automated deployment and customization of routing overlays using resources from heterogeneous experimentation platforms. The main contribution of this work is to enable researchers to easily integrate different resources, such as simulated, emulated or physical nodes, on a same experiment, using a network overlay, thus addressing one of the main concerns previously mentioned.

We started by adding support to easily build routing overlays on PlanetLab, and providing the ability to customize network traffic by adding user defined filters to packets traversing the overlay tunnels [48]. We then improved this work by adding the ability to include simulated nodes from the ns-3 backend and emulated nodes from the linux containers backend into a single overlay network. We demonstrated the use of NEPI to build adn control routing overlays which incorporate resources from different on the ns-3 2012 community workshop [74].

- **Content Centric Networks Live Experimentation**

  Realistic experimentation on top of Internet-like environments is key to evaluate the feasibility of world wide deployment of CCNx, and to assess the impact of existing Internet traffic conditions on CCN traffic. However, deploying live experiments on the Internet is a difficult and error prone task, specially when performed manually.

  To address this issue, during the last year, we extended NEPI, a framework for managing network experiments, to support easy design, and automated deployment and control, of CCNx experiments on the PlanetLab testbed. Among other features, NEPI now enables the deployment of user modified CCNx sources on arbitrary PlanetLab nodes, and the creation of tunnels to enable the use of multicast FIB entries between CCNx daemons over the Internet. By supporting easy CCNx experimentation on PlanetLab, NEPI can help to explore the co-existence of CCN and TCP/IP architecture.

  This work was presented as a poster and a demo at CCNxCon 2012, the CCNx http://www.ccnx.org/ community meeting [73]. The work had a very good reception and gained NEPI some new users.

  An online tutorial and demo were also made available at NEPI's web page http://nepi.inria.fr/wiki/nepi/CCNxOnPlanetLabEurope, for dissemination purposes.

- **Smooth-transition: a new methodology for dealing with various network experiment environments**

  The smooth-transition is a new methodology, which supports various network experiment environments covering from pure simulation through realistic emulation consistently. The reproducibility in experimental network research is getting important feature for iterative experiments in short-term and long-term period. The main idea of this concept is providing the reproducibility in a broader sense. So far, we had to implement different experiments by different environment, such as simulation, application-level emulation, and link-level emulation. Whereas the smooth-transition is able to keep the context of the experiments started from a pure simulation up to a realistic emulation gradually. That means the user does not need to waste time any more for learning and following a lot of documents and manuals from each different environment. Moreover, anyone can easily start to use the testbed and to develop inside (i.e. protocol stack). Because NS3 which is the most popular and powerful network simulator has been used in this concept as an experiment engine.

  The smooth-transition employees Network Experiment Programming Interface (NEPI) to conduct all functions, such as composing scenario, node deployment, experiment control, and resource management. The core of building this concept is NS3 which has Emulation (EMU) and Direct Code Execution (DCE) modules. EMU supports to use real network devices instead of NS3 MAC and PHY layer implementations. DCE is able to launch real application on top of NS3 protocol stacks. Furthermore, real Linux kernel (currently, net-next 2.6 is available) can replace NS3 Internet protocols by its advanced mode. This concept needs back-end system covering all experiment nodes. Control and Management Framework (OMF) plays an important role as a software framework to control and manage an wireless network testbed, and all messages are exchanged by Extensible Messaging and Presence Protocol (XMPP). Nitos scheduler has been adopted as a reservation system http://nitlab.inf.uth.gr/NITlab/index.php/scheduler. The user can reserve a time slot, nodes, and wireless channels through its web page. In addition, SFA supports that the testbed is federated with other ones of outside.

The testbed provides PCAP files as a common outcome, and this file contains captured in and out packets. However, the file size is easily over gigabytes, then it makes a very long delay to process dozens of that files. To reduce the processing time efficiently, we are using an indexing scheme for fast collecting desired packets by filtering. In particular, this scheme is very useful to find packets occurred rarely, when an detailed analysis is required for an network event, such as retransmission, intrusion detection, and node association/disassociation. The indexing information is stored in a database file, and it does not need to be modified after making the file. The size of the file is very small compared with the PCAP file, so it provides fast packet filtering permanently, even after leaving the testbed. This work, post-processing of PCAP files, is in a collaboration with Diego Dujovne and Luciano Ahumada from the Universidad Diego Portales of Chili. Especially, YoungHwan Kim, a postdoc of the Planète group, has been currently dispatched for this collaboration for fourteen weeks (September 15 2012 $\sim$ January 26 2013) in Santiago, Chile.

- **The FIT experimental platform**

  We have started, since 2011, the procedure of building a new experimental platform at Sophia-Antipolis, in the context of the FIT Equipment of Excellence project. This platform has two main goals : the first one is to enable highly controllable experiments due to its anechoic environment. These experiments can be either hybrid-experiments (as NEPI will be deployed) or federated-experiments through several testbeds. The second goal is to make resource consuming experiments (like CCNx) possible due to some powerful servers that will be installed and connected to the PlanetLab testbed. During 2012, the specifications has been defined and the procedure will continue during the next year.

- **Network Simulations on a Grid**

  We studied an hybrid approach for the evaluation of networking protocols based on the ns-3 network simulator and a Grid testbed. We analyzed the performance of the approach using a simple use case. Our evaluation shows that the scalability of our approach is mainly limited by the processor speed and memory capacities of the simulation node. We showed that by exploiting the emulation capacity of ns-3, it is possible to map complex network scenarios on grid nodes. We also proposed a basic mapping algorithm to distribute a network scenario on several node [32].

RAP Project-Team

4. New Results

4.1. Algorithms: Bandwidth Allocation in Optical Networks

Participants: Christine Fricker, Philippe Robert, James Roberts.

The development of dynamic optical switching is widely recognized as an essential requirement to meet anticipated growth in Internet traffic. Since September 2009, RAP has investigated the traffic management and performance evaluation issues that are particular to this technology. A first analysis of passive optical networks used for high speed Internet access led to the proposal of an original dynamic bandwidth allocation algorithm and to an evaluation of its traffic capacity. Our activity on optical networking is carried out in collaboration with Orange Labs with whom we have a research contract. We have also established contacts with Alcatel-Lucent Bell Labs and had fruitful exchanges with Iraj Saniee and his team on their proposed time-domain wavelength interleaved networking architecture (TWIN).

We have analyzed the traffic capacity of wavelength division multiplexing (WDM), passive optical networks (PONs) where user stations (optical network units) are equipped with tunable transmitters. For these systems users can use any of the multiple wavelengths to transmit their data but only within the limit determined by the number of transmitters they possess. A mean field approximation is used to estimate the capacity of a limited-gated multiserver polling system with a limit on the number of servers a given station can use simultaneously. The approximation provides an expression for the stability limit under very general assumptions about the traffic process and system configuration.

In 2011, we began work on bandwidth allocation in meshed networks. We have evaluated the TWIN architecture in a metropolitan area network with an original medium access control (MAC) algorithm. This algorithm was inspired by our prior work on access networks and ensures an efficient and fair allocation of bandwidth to flows between network nodes.

The TWIN architecture is not extensible to a wide area for reasons of scalability and the excessive signalling delay between geographically distant nodes. We have therefore invented a new notion of a multipoint-to-multipoint lightpath that avoids these problems. A patent relating to this invention has been granted. This patent is owned by Orange following the terms of our contract with them. The paper [16] describes the invention and its evaluation. A major advantage demonstrated in this paper is the energy saving achieved by the use of the proposed optical technology in place of electronic routers. An extended version of the paper has been accepted for publication in Journal of Optical Communication and Networking [24].

Ongoing research seeks to apply this type of networking solution to data centres, on one hand, and to geographically spread tier-1 Internet carrier networks, on the other. Some of this work is performed in collaboration with Orange Labs under the terms of our research contract. An interesting new development is the application of new coherent optical technology that allows tunable receivers as well as tunable transmitters. We are evaluating the performance of a bandwidth allocation algorithm that exploits this technology.

A wider reaching collaboration has been established under the terms of a Celtic Plus project called SASER. This project was approved by the EU in 2012 and funding has been obtained for our participation from the French authorities. The project kickoff meeting was held in November 2012. Our contribution relates to the use of TWIN to create an extended metropolitan optical network. Our partners in the corresponding work package task are Orange, Telecom Bretagne and the engineering school ENSSAT. Overall reponsibility for the work package (where alternative optical network architectures are also evaluated) is with Alcatel-Lucent Bell Labs.

4.2. Algorithms: Content-Centric Networking

Participants: Mathieu Feuillet, Christine Fricker, Philippe Robert, James Roberts, Nada Sbihi.

RAP is participating in an ANR project named CONNECT which contributes to the definition and evaluation of a new paradigm for the future Internet: a content-centric network (CCN) where, rather than interconnecting remote hosts like IP, the network directly manages the information objects that users publish, retrieve and exchange. CCN has been proposed by Van Jacobson and colleagues at the Palo Alto Research Center (PARC). In CCN, content is divided into packet-size chunks identified by a unique name with a particular hierarchical structure. The name and content can be cryptographically encoded and signed, providing a range of security levels. Packets in CCN carry names rather than addresses and this has a fundamental impact on the way the network works. Security concerns are addressed at the content level, relaxing requirements on hosts and the network. Users no longer need a universally known address, greatly facilitating management of mobility and intermittent connectivity. Content is supplied under receiver control, limiting scope for denial of service attacks and similar abuse. Since chunks are self-certifying, they can be freely replicated, facilitating caching and bringing significant bandwidth economies. CCN applies to both stored content and to content that is dynamically generated, as in a telephone conversation, for example. RAP is contributing to the design of CCN in two main areas:

- the design and evaluation of traffic controls, recognizing that TCP is no longer applicable and queue management will require new, name-based criteria to ensure fairness and to realize service differentiation;
- the design and evaluation of replication and caching strategies that realize an optimal trade-off of expensive bandwidth for cheap memory.

The team also contributes to the development of efficient forwarding strategies and the elaboration of economic arguments that make CCN a viable replacement for IP. CONNECT partners are Alcatel-Lucent (lead), Orange, Inria/RAP, Inria/PLANETE, Telecom ParisTech, UPMC/LIP6.

A paper describing a proposed flow-aware approach for CCN traffic management and its performance evaluation has been presented at the conference Infocom 2012 [20]. We have reviewed the literature on cache performance (dating from early work on computer memory management) and identified a practical and versatile tool for evaluating the hit rate (proportion of requests that are satisfied from the cache) as a function of cache size and the assumed object popularity law. This approximate method was first proposed in 2002 by Che, Tung and Wang for their work on web caching. We applied this approximation to evaluate CCN caching performance taking into account the huge population and diverse popularity characteristics that make other approaches ineffective [19]. The excellent accuracy of this method over a wide range of practically relevant traffic models has been explained mathematically [18]. CONNECT ends in December 2012. We are currently defining a new project proposal that should be submitted to the ANR INFRA call in February 2013.

4.3. Scaling Methods: Fluid Limits in Wireless Networks

Participant: Philippe Robert.

This is a collaboration with Amandine Veber (CMAP, École Polytechnique). The goal is to investigate the stability properties of wireless networks when the bandwidth allocated to a node is proportional to a function of its backlog: if a node of this network has x requests to transmit, then it receives a fraction of the capacity proportional to $\log(1 + x)$, the logarithm of its current load. A fluid scaling analysis of such a network is presented. We have shown that the interaction of several time scales plays an important role in the evolution of such a system, in particular its coordinates may live on very different time and space scales. As a consequence, the associated stochastic processes turn out to have unusual scaling behaviors which give an interesting fairness property to this class of algorithms. A heavy traffic limit theorem for the invariant distribution has also been proved. A generalization to the resource sharing algorithm for which the $\log$ function is replaced by an increasing function.

4.4. Algorithms: Distributed Hash Tables

Participants: Mathieu Feuillet, Philippe Robert.

The Distributed Hash Table (DHTs) consists of a large set of nodes connected through the Internet. Each file contained in the DHT is stored in a small subset of these nodes. Each node breaks down periodically and it is necessary to have back-up mechanisms in order to avoid data loss. A trade-off is necessary between the bandwidth and the memory used for this back-up mechanism and the data loss rate. Back-up mechanisms already exist and have been studied thanks to simulation. To our knowledge, no theoretical study exists on this topic. We modeled this problem thanks to standard queues in order to understand the behavior of a single file and the global dynamic of the system. With a very simple centralized model, we have been able to emphasise a trade-off between capacity and life-time with respect to the duplication rate. From a mathematical point of view, we have been able to study different time scales of the system with an averaging phenomenon. A paper has been submitted on this subject for the case where there are at most two copies of each file [25]. An article for the general case is in preparation. A more sophisticated distributed model with mean field techniques is under investigation.

On the side of this project, we notably studied the distribution of hitting times of the classical Ehrenfest and Engset models by using martingale techniques, furthermore their asymptotic behavior has been analyzed when the size of the system increases to infinity [11].

4.5. Stochastic Modeling of Biological Networks

Participants: Emanuele Leoncini, Philippe Robert.

This is a collaboration with Vincent Fromion from INRA Jouy en Josas, which started on October 2010.

The goal is to propose a mathematical model of the production of proteins in prokaryotes. Proteins are biochemical compounds that play a key role in almost all the cell functions and are crucial for cell survival and for life in general. In bacteria the protein production system has to be capable to produce abut 2500 different types of proteins in different proportions (from few dozens for the replication machinery up to 100000 for certain key metabolic enzymes). Bacteria uses more than the 85% of their resources to the protein production, making it the most relevant process in these organisms. Moreover this production system must meet two opposing problems: on one side it must provide a minimal quantity for each protein type in order to ensure the smooth-running of the cell, on the other side an “overproduction policy” for all the proteins is infeasible, since this would impact the global performance of the system and of the bacterium itself.

Gene expression is intrinsically a stochastic process: gene activation/deactivation occurs by means the encounter of polymerase/repressor with the specific gene, moreover many molecules that take part in the protein production act at extremely low concentrations. We have restated mathematically the classical model using Poisson point processes (PPP). This representation, well-known in the field of queueing networks but, as far as we know, new in the gene expression modeling, allowed us to weaken few hypothesis of the existing models, in particular the Poisson hypothesis, which is well-suited in some cases, but that, in some situations, is far from the biological reality as we consider for instance the protein assemblage. See [12].

The theoretical environment of Poisson point processes has lead us to propose a new model of gene expression which captures on one side the main mechanisms of the gene expression and on the other side it tries to consider hypothesis that are more significant from a biological viewpoint. In particular we have modeled: gene activation/deactivation, mRNA production and degradation, ribosome attachment on mRNA, protein elongation and degradation. We have shown how the probability distribution of the protein production and the protein lifetime may have a significant impact on the fluctuations of the number of proteins. We have obtained analytic formulas when the duration of protein assemblage and degradation follows a general probability distribution, i.e. without the Poisson hypothesis. In particular, by using a PPP representation we have been able to include the deterministic continuous phenomenon of protein degradation, which is the main protein degradation mechanism for stable proteins. We have showed moreover that this more realistic description is surprisingly identical in distribution with the classic assumption of protein degradation by means of a degrading protein (*proteosome*). We have used our model also to compare the variances resulting by choosing different hypotheses for the probability elongation, in particular we have hypothesize the protein assembly to be deterministic. This assumption is justified because of the elongation step, which consists of a large number of elementary steps, can be described by the sum of exponential steps and the resulting distribution is well approximated by a Gaussian distribution because of the central limit theorem. Under the hypothesis of small variance of the resulting Gaussian distribution, we can assume the elongation step to be deterministic. The model has showed how, under the previous hypothesis, the variance on the number of proteins is bigger than the classical model with the Poisson hypothesis.

We have developed a C++ stochastic simulator for our general model, which has allowed the computation of variance when it was not possible to derive explicit analytic close formulas and the simulation of some extension of the actual model.

4.6. Stochastic Networks: Large Bike Sharing Systems

Participants: Christine Fricker, Hanene Mohamed, Danielle Tibi.

This is a collaboration with Nicolas Gast (EPFL) starting in December 2010. Bike sharing systems were launched by numerous cities as a part of urban transportation, for example Velib in 2007 (20 000 bikes, 1 500 stations). One of the major issues is the availability of the resources: bikes or free slots. These systems become a hot topic in Operation Research but studies on these stochastic networks are very few. To our knowledge, no theoretical study of such bike sharing systems exist taking into account the limited capacity of the stations.

We modeled this system in a symmetric case. Mean field limit gives the dynamic of a large system and the limiting stationary behavior of a single station as the system gets large. Analytical results are obtained and convergence proved in the standard model via Lyapounov functions. It allows to find the best ratio of bikes par station and to measure the improvement of incentive mechanisms, as choosing among two stations for example. Redistribution by trucks is also investigated. See [26].

Further results have been obtained for some heterogeneous systems. By mean field techniques, analytical results are obtained with Hanene Mohamed for systems with clusters (see [17]).

In a work in progress with *Danielle Tibi* , a more direct method is used when the network has a product form invariant measure by central and local limit theorem. It is a way to prove in this case the equivalence of ensembles, known in physic statistics. It applies to the simplest non homogeneous model. It gives a way to generalize the cluster case.

4.7. Random Graphs

Participant: Nicolas Broutin.

4.7.1. Connectivity in models of wireless networks

This is joint work with S. Boucheron (Paris 7), L. Devroye (McGill), N. Fraiman (McGill), and G. Lugosi (Pompeu Fabra).

The traditional models for wireless networks rely on geometric random graphs. However, if one wants to ensure that the graph be fully connected the radius of influence (hence the power necessary, and number of links) is too large to be fully scalable. Recently some models have been proposed that skim the neighbours and only retain a random subset for each node, hence creating a sparser overlay that would hopefully be more scalable. The first results on the size of the subsets which guarantee connectivity of overlay (the irrigation graph) [3] confirm that the average number of links per node is much smaller, but it remains large. These results motivate further investigations on the size of the largest connected component when one enforces a constant average degree which are in the process of being written.

4.7.2. Random graphs and minimum spanning trees

This is a long term collaboration with L. Addario-Berry (McGill), C. Goldschmidt (Oxford) and G. Miermont (ENS Lyon).

The random graph of Erdős and Rényi is one of the most studied models of random networks. Among the different ranges of density of edges, the “critical window” is the most interesting, both for its applications to the physics of phase transitions and its applications to combinatorial optimization (minimum spanning tree, constraint satisfaction problems). One of the major questions consists in determining the distribution of distances between the nodes. A limit object (a scaling limit) has been identified, that allows to describe precisely the first order asymptotics of pairwise distances between the nodes. This limit object is a random metric space whose definition allows to exhibit a strong connection between random graphs and the continuum random tree of Aldous. A variety of questions like the diameter, the size of cycles, etc, may be answered immediately by reading them on the limit metric space [2].

In a stochastic context, the minimum spanning tree is tightly connected to random graphs via Kruskal's algorithm. Random minimum spanning trees have attracted much research because of their importance in combinatorial optimization and statistical physics; however, until now, only parameters that can be grasped by local arguments had been studied. The scaling limit of the random graphs obtained in [2] permits to describe precisely the metric space scaling limit of a random minimum spanning tree [21], which identifies a novel continuum random tree which is truly different from that of Aldous.

4.7.3. Analysis of recursive partitions

This is joint work with R. Neininger (Frankfurt) and H. Sulzbach (Frankfurt/McGill).

The quadtrees are essential data structures that permit to store and manipulate geometric data by building a recursive partition of the space. In order to evaluate their performance, Flajolet and his co-authors have estimated the average cost of reporting all the data matching certain random queries. When the query does not fully specify all the fields, one talk about a partial match query. Such queries are ubiquitous, but analyzing their behaviour turns out to be intricate, and no performance guarantee was available in the form of a bound on the probability that any query would take much more time that one expects. [14] provides such guarantees by analysing the behaviour of all the queries at the same time, as a process. This yields estimates for the cost of the worst possible query (not a uniformly random one), as well as asymptotics for the variance and higher moments.

This line of research has motivated the analysis of the related combinatorial model of recursive lamination of the disk. The model had been recently introduced, but no full analysis was available. The techniques developed in the context of quadtrees have inspired a proof that the dual tree of the recursive lamination does converge to a limit tree-like metric space which is identified [23].

4.7.4. Navigation and point location in Poisson Delaunay triangulation

Nicolas Broutin has recently initiated a project with O. Devillers (Inria Sophia) and R. Hemsley (Inria Sophia) concerning the performance of local routing algorithms in plane subdivisions. Such algorithms also turn out to be important for the *point location* problem: for instance, finding the face of the subdivision which contains a query point is the first step towards inserting this point as a vertex. The aim is to prove that when the subdivision consists of the faces of a Delaunay triangulation, and when the points are random, any natural strategy which would take you closer to the aim performs well. Preliminary results about a specific routing algorithm, the cone walk, that we designed for its amenability to analysis appear in [22].

4.8. Stochastic Networks: Jackson Networks

Participant: Danielle Tibi.

Lyapounov functions and essential spectral radius of Jackson networks, joint work with I. Ignatiouk-Robert (University of Cergy-Pontoise). A family of explicit multiplicative Lyapounov functions is constructed for any stable Jackson network. Optimizing the multiplicative factor over this family provides an upper bound for the essential spectral radius of the associated Markov process. For some particular classes of Jackson networks, this upper bound coincides with a lower bound derived from large deviations arguments, thus providing the exact value of the essential spectral radius. The main example is given by Jackson networks with routing matrix having a tree structure (in the sense that for any node i, at most one other node can route its customers to i).

The result also holds for other types of routing matrices (e.g. completely symmetrical), under some conditions over the different arrival and service rates. See [27].

REGAL Project-Team

6. New Results

6.1. Introduction

In 2012, we focused our research on the following areas:

- *Management of distributed data.*
- *Performance and robustness of Systems Software in multicore architectures.*

6.2. Distributed algorithms for dynamic networks

Participants: Luciana Arantes [correspondent], Olivier Marin, Sébastien Monnet, Franck Petit [correspondent], Maria Potop-Butucaru, Pierre Sens, Julien Sopena, Raluca Diaconu, Ruijing Hu, Anissa Lamani, Sergey Legtchenko, Jonathan Lejeune, Karine Pires, Guthemberg Silvestre, Véronique Simon.

This objective aims to design distributed algorithms adapted to new large scale or dynamic distributed systems, such as mobile networks, sensor networks, P2P systems, Grids, Cloud environments, and robot networks. Efficiency in such demanding environments requires specialised protocols, providing features such as fault or heterogeneity tolerance, scalability, quality of service, and self-stabilization. Our approach covers the whole spectrum from theory to experimentation. We design algorithms, prove them correct, implement them, and evaluate them in simulation, using OMNeT++ or PeerSim, and on large-scale real platforms such as Grid'5000. The theory ensures that our solutions are correct and whenever possible optimal; experimental evidence is necessary to show that they are relevant and practical.

Within this thread, we have considered a number of specific applications, including massively multi-player on-line games (MMOGs) and peer certification.

Since 2008, we have obtained results both on fundamental aspects of distributed algorithms and on specific emerging large-scale applications.

We study various key topics of distributed algorithms: mutual exclusion, failure detection, data dissemination and data finding in large scale systems, self-stabilization and self-* services.

6.2.1. Mutual Exclusion and Failure Detection.

Mutual Exclusion and Fault Tolerance are two major basic building blocks in the design of distributed systems. Most of the current mutual exclusion algorithms are not suitable for modern distributed architectures because they are not scalable, they ignore the network topology, and they do not consider application quality of service constraints. Under the ANR Project *MyCloud* and the FSE *Nu@age*, we study locking algorithms fulfilling some QoS constraints often found in Cloud Computing [38].

A classical way for a distributed system to tolerate failures is to detect them and then recover. It is now well recognized that the dominant factor in system unavailability lies in the failure detection phase. Regal has worked for many years on practical and theoretical aspects of failure detections and pioneered hierarchical scalable failure detectors. [2] Since 2008, we have studied the adaptation of failure detectors to dynamic networks. Following the model introduced in [18], we have proposed new algorithms to detect crashes and Byzantine behaviors [32].

These algorithms were designed as part of the ANR Project SHAMAN.

[2]Recent work by Leners et al published in SOSP 2011 uses our DSN 2003 paper as basis for performance comparison

6.2.2. Self-Stabilization and Self-* Services.

We have also approached fault tolerance through self-stabilization. Self-stabilization is a versatile technique to design distributed algorithms that withstand transient faults. In particular, we have worked on the unison problem, [3] i.e., the design of self-stabilizing algorithms to synchronize a distributed clock. As part of the ANR project *SPADES*, we have proposed several snap-stabilizing algorithms for the message forwarding problem that are optimal in terms of number of required buffers [36]. A snap-stabilizing algorithm is a self-stabilizing algorithm that stabilizes in 0 steps; in other words, such an algorithm always behaves according to its specification.

Finally, we have applied our expertise in distributed algorithms for dynamic and self-* systems in domains that at first glance seem quite far from the core expertise of the team, namely ad-hoc systems and swarms of mobile robots. In the latter, as part of ANR project *R-Discover*, we have studied various problems such as exploration [29], and gathering [15].

6.2.3. Dissemination and Data Finding in Large Scale Systems.

In the area of large-scale P2P networks, we have studied the problems of data dissemination and overlay maintenance, i.e., maintenance of a logical network built over the a P2P network. First, we have proposed efficient distributed algorithms to ensure data dissemination to a large set of nodes. Also, we have introduced a new method to compare dissemination algorithms over various topologies [35].

6.2.4. MMOGs.

Peer-to-peer overlay networks can be used to build scalable infrastructures for MMOGs. Our work on MMOGs has primarily focused on the impact of latency constraints in dynamic distributed systems. In online P2P games, players are connected by a logical graph, implemented as an overlay network. Latency constraints imply that players that interact must remain close in the overlay, even when the mobility of players induces rapid changes in the graph.

We have also addressed problems related to cheating and arbitration. In a distributed system, certification of entities makes it possible to circumscribe malicious behavior, such as cheating in games. Certification requires the use of a trusted third party and is traditionally done centrally. At a large scale, however, centralized certification represents a bottleneck and a single point of attack or failure. We have proposed solutions based on distributed reputations to identify trusted nodes and use them as game referees to detect and prevent cheating [46]. Our method relies on previous work on the subject of trusted node collaboration to ensure reliable distributed certification [4].

6.3. Management of distributed data

Participants: Mesaac Makpangou, Olivier Marin, Sébastien Monnet, Pierre Sens, Marc Shapiro, Julien Sopena, Gaël Thomas, Pierpaolo Cincilla, Raluca Diaconu, Sergey Legtchenko, Jonathan Lejeune, Karine Pires, Thomas Preud homme, Masoud Saeida Ardekani, Guthemberg Silvestre, Pierre Sutra, Marek Zawirski, Annette Bieniusa, Pierpaolo Cincilla, Véronique Simon, Mathieu Valero.

Sharing information is one of the major reasons for the use of large-scale distributed computer systems. Replicating data at multiple locations ensures that the information persists despite the occurrence of faults, and improves application performance by bringing data close to its point of use, enabling parallel reads, and balancing load. This raises numerous issues: where to store or replicate the data, in order to ensure that it is available quickly and remains persistent despite failures and disconnections; how to ensure consistency between replicas; when and how to move data to computation, or computation to data, in order to improve response time while minimizing storage or energy usage; etc. The Regal group works on several key issues related to replication:

[3] C. Boulinier, F. Petit, and V. Villain. Synchronous vs. asynchronous unison. Algorithmica, 51(1):61-80, 2008

[4] Erika Rosas, Olivier Marin and Xavier Bonnaire. CORPS: Building a Community Of Reputable PeerS in Distributed Hash Tables. The Computer Journal, 54(10):1721-1735(2011)

- Replica placement for fault tolerance and latency in the presence of churn,
- scalable strong consistency for replicated databases, and
- theory and practice of eventual consistency.

6.3.1. Distributed hash tables

A DHTs replicates data and spreads the replicas uniformly across a large number of nodes. Being very scalable and fault-tolerant, DHTs are a key component for dependable and secure applications, such as backup systems, distributed file systems, multi-range query systems, and content distribution systems.

Despite the advantages of DHTs, several studies show that they become inefficient in environments subject to churn, i.e., with many node arrivals and departures. We therefore propose a new replication mechanism for DHTs that is churn resilient [20]. RelaxDHT relaxes placement constraints, in order to avoid redundant data transfers and to increase parallelism. RelaxDHT loses up to 50% fewer data blocks that the well-known PAST DHT.

6.3.2. Strong consistency

When data is updated somewhere on the network, it may become inconsistent with data elsewhere, especially in the presence of concurrent updates, network failures, and hardware or software crashes. A primitive such as consensus (or equivalently, total-order broadcast) synchronises all the network nodes, ensuring that they all observe the same updates in the same order, thus ensuring strong consistency. However the latency of consensus is very large in wide-area networks, directly impacting the response time of every update. Our contributions consist mainly of leveraging application-specific knowledge to decrease the amount of synchronisation.

To reduce the latency of consensus, we study *Generalised Consensus* algorithms, i.e., ones that leverage the commutativity of operations or the spontaneous ordering of messages by the network. We propose a novel protocol for generalised consensus that is optimal, both in message complexity and in faults tolerated, and that switches optimally between its fast path (which avoids ordering commuting requests) and its classical path (which generates a total order). Experimental evaluation shows that our algorithm is much more efficient and scales better than competing protocols.

When a database is very large, it pays off to replicate only a subset at any given node; this is known as partial replication. This allows non-overlapping transactions to proceed in parallel at different locations and decreases the overall network traffic. However, this makes it much harder to maintain consistency. We designed and implemented two *genuine* consensus protocols for partial replication, i.e., ones in which only relevant replicas participate in the commit of a transaction.

Another research direction leverages isolation levels, particularly Snapshot Isolation (SI), in order to parallelize non-conflicting transactions on databases. We prove a novel impossibility result, namely that a system cannot have both genuine partial replication and SI. We designed an efficient protocol that maintains the most important features of SI, but side-steps this impossibility. Finally, we study the trade-offs between freshness (and hence low abort rates) and space complexity in computing snapshots, as required by SI and its variants.

Parallel transactions in distributed DBs incur high overhead for concurrency control and aborts. Our Gargamel system proposes an alternative approach by pre-serializing possibly conflicting transactions, and parallelizing non-conflicting update transactions to different replicas. It system provides strong transactional guarantees. In effect, Gargamel partitions the database dynamically according to the update workload. Each database replica runs sequentially, at full bandwidth; mutual synchronisation between replicas remains minimal. Our simulations show that Gargamel improves both response time and load by an order of magnitude when contention is high (highly loaded system with bounded resources), and that otherwise slow-down is negligible. This is published at ICPADS 2012 [27].

Our current experiments aim to compare the practical pros and cons of different approaches to designing large-scale replicated databases, by implementing and benchmarking a number of different protocols.

Our study the trade-offs between freshness and meta-date overhead, is published in HotCDP 2012 [43].

6.3.3. Eventual consistency

Eventual Consistency (EC) aims to minimize synchronisation, by weakening the consistency model. The idea is to allow updates at different nodes to proceed without any synchronisation, and to propagate the updates asynchronously, in the hope that replicas converge once all nodes have received all updates. EC was invented for mobile/disconnected computing, where communication is impossible (or prohibitively costly). EC also appears very appealing in large-scale computing environments such as P2P and cloud computing. However, its apparent simplicity is deceptive; in particular, the general EC model exposes tentative values, conflict resolution, and rollback to applications and users. Our research aims to better understand EC and to make it more accessible to developers.

We propose a new model, called *Strong Eventual Consistency* (SEC), which adds the guarantee that every update is durable and the application never observes a roll-back. SEC is ensured if all concurrent updates have a deterministic outcome. As a realization of SEC, we have also proposed the concept of a Conflict-free Replicated Data Type (CRDT). CRDTs represent a sweet spot in consistency design: they support concurrent updates, they ensure availability and fault tolerance, and they are scalable; yet they provide simple and understandable consistency guarantees.

This new model is suited to large-scale systems, such as P2P or cloud computing. For instance, we propose a “sequence” CRDT type called Treedoc that supports concurrent text editing at a large scale, e.g., for a wikipedia-style concurrent editing application. We designed a number of CRDTs such as counters (supporting concurrent increments and decrements), sets (adding and removing elements), graphs (adding and removing vertices and edges), and maps (adding, removing, and setting key-value pairs). In particular, we publish a study of the concurrency semantics of sets in DISC 2012 [48], [22].

On the theoretical side, we identified sufficient correctness conditions for CRDTs, viz., that concurrent updates commute, or that the state is a monotonic semi-lattice. CRDTs raise challenging research issues: What is the power of CRDTs? Are the sufficient conditions necessary? How to engineer interesting data types to be CRDTs? How to garbage collect obsolete state without synchronisation, and without violating the monotonic semi-lattice requirement?

We are currently developing a very large-scale CRDT platform called SwiftCloud, which aims to scale to millions of clients, deployed inside and outside the cloud.

6.4. Improving the Performance and Robustness of Systems Software in Multicore Architectures

6.4.1. Managed Runtime Environments

Participants: Bertil Folliot, Julia Lawall, Gilles Muller [correspondent], Marc Shapiro, Julien Sopena, Gaël Thomas, Florian David, Lokesh Gidra, Jean-Pierre Lozi, Thomas Preud homme, Suman Saha, Harris Bakiras, Arie Middelkoop, Koutheir Attouchi.

Today, multicore architectures are becoming ubiquitous, found even in embedded systems, and thus it is essential that managed languages can scale on multicore processors. We have found that a major scalability bottleneck is the implementation of high contention locks, which can overload the bus, eliminating all performance benefits from adding more cores. To address this issue, as part of the PhD of Jean-Pierre Lozi, we have developed remote core locking (RCL), in which highly contended locks are implemented on a dedicated server, minimizing bus traffic and improving application scalability (USENIX ATC 2012 [24]). This work initially targeted C code but is now being adapted to the needs of Java applications in the PhD of Florian David. Another bottleneck in the support for managed languages is the garbage collector. As part of the PhD of Lokesh Gidra, we have identified the main sources of overhead.

6.4.2. Systems software robustness

A new area of research for Regal, with the arrival of Gilles Muller in 2009 as Inria Senior Research Scientist and Julia Lawall in 2011 as Inria Senior Research Scientist, is on improving the reliability of operating systems code. Muller and Lawall previously developed Coccinelle, a scriptable program matching and transformation tool for C code that is now commonly used in the open-source development community, including by the developers of Linux, Wine and Dragonfly BSD. Based on Coccinelle, we have developed a new approach to inferring API function usage protocols from software, relying on knowledge of common code structures (Software – Practice and Experience [19]).

We have also proposed a method for automatically identifying bug-fixing patches, with the goal of helping developers maintain stable versions of the software (ICSE 2012 [45]) and have designed an approach to automatically generating a robust interface to the Linux kernel, to provide developers of new kernel-level code more feedback in the case of a misunderstanding of kernel API usage conventions (ASE 2012 [24]).

RMOD Project-Team

5. New Results

5.1. Object serializer

Participants: Martin Dias [Correspondant], Mariano Martinez-Peck, Stéphane Ducasse.

Fuel: A Fast General Purpose Object Graph Serializer Since objects need to be stored and reloaded on different environments, serializing object graphs is a very important activity. There is a plethora of serialization frameworks with different requirements and design trade-offs. Most of them are based on recursive parsing of the object graphs, an approach which often is too slow. In addition, most of them prioritize a language-agnostic format instead of speed and language-specific object serialization. For the same reason, such serializers usually do not support features like class-shape changes, global references or executing pre and post load actions. Looking for speed, some frameworks are partially implemented at Virtual Machine (VM) level, hampering code portability and making them difficult to understand, maintain and extend. That is why we work on Fuel, a general-purpose object serializer based on these principles: (1) speed, through a compact binary format and a pickling algorithm which invests time in serialization for obtaining the best performance on materialization; (2) good object-oriented design, without special help at VM; (3) serialize any object, thus have a full-featured language-specific format. We implement and validate this approach in Pharo, where we demonstrate that Fuel is faster than other serializers, even those with special VM support. The extensibility of Fuel made possible to successfully serialize various objects: classes in Newspeak, debugger stacks, and full CMS object graphs [11].

5.2. Cycles and dependencies

Participants: Stéphane Ducasse [Correspondant], Nicolas Anquetil, Muhammad Bhatti.

OZONE: Layer Identification in the presence of Cyclic Dependencies A layered software architecture helps understanding the role of software entities (e.g., packages or classes) in a system and hence, the impact of changes on these entities. However, the computation of an optimal layered organization in the presence of cyclic dependencies is difficult. We develop an approach that (i) provides a strategy supporting the automated detection of cyclic dependencies, (ii) proposes heuristics to break cyclic dependencies, and (iii) computes an organization of software entities in multiple layers even in presence of cyclic dependencies. Our approach performs better than the other existing approaches in terms of accuracy and interactivity, it supports human inputs and constraints. We compare this approach to existing solutions and apply it on two large software systems to identify package layers. The results are manually validated by software engineers of the two systems [12].

Efficient Retrieval and Ranking of Undesired Package Cycles in Large Software Systems Many design guidelines state that a software system architecture should avoid cycles between its packages. Yet such cycles appear again and again in many programs. We believe that the existing approaches for cycle detection are too coarse to assist developers to remove cycles from their programs. We design an efficient algorithm that performs a fine-grained analysis of cycles among application packages. In addition, we define multiple metrics to rank cycles by their level of undesirability, prioritizing cycles that are the more undesired by developers. We compare these multiple ranking metrics on four large and mature software systems in Java and Smalltalk [14].

Resolving cyclic dependencies between packages with Enriched Dependency Structural Matrix Dependency Structural Matrix (DSM) is an approach originally developed for process optimization. It has been successfully applied to identify software dependencies among packages and subsystems. A number of algorithms have been proposed to compute the matrix so that it highlights patterns and problematic dependencies between subsystems. However, existing DSM implementations often miss important information to fully support reengineering effort. For example, they do not clearly qualify and quantify problematic relationships, information that is crucial to support remediation tasks. We propose Enriched Dependency Structural Matrix (eDSM), which provides small multiple views and micro-macro readings by adding fine-grained information in each cell of the matrix. Each cell is enriched with contextual information about (i) the type of dependencies (inheritance, class reference, etc.), (ii) the proportion of referencing entities, (iii) the proportion of referenced entities. We distinguish independent cycles and stress potentially simple fixes for cycles using coloring information. This work is language independent and has been implemented on top of the Moose software analysis platform.We improved the cell content information view based on user feedback and performed multiple validations: two different case studies on Moose and Seaside software; one user study for validating eDSM as a usable approach for developers. Solutions to problems identified with eDSM have been performed and retrofitted in analyzed software [13].

5.3. Warnings and bugs

Participants: Simon Allier [Correspondant], Andre Hora, Nicolas Anquetil, Muhammad Bhatti, Stéphane Ducasse.

A Framework to Compare Alert Ranking Algorithms To improve software quality, rule checkers statically check if a software contains violations of good programming practices. On a real sized system, the alerts (rule violations detected by the tool) may be numbered by the thousands. Unfortunately, these tools generate a high proportion of "false alerts", which in the context of a specific software, should not be fixed. Huge numbers of false alerts may render impossible the finding and correction of "true alerts" and dissuade developers from using these tools. In order to overcome this problem, the literature provides different ranking methods that aim at computing the probability of an alert being a "true one". We propose a framework for comparing these ranking algorithms and identify the best approach to rank alerts. We have selected six algorithms described in literature. For comparison, we use a benchmark covering two programming languages (Java and Smalltalk) and three rule checkers (FindBug, PMD, SmallLint). Results show that the best ranking methods are based on the history of past alerts and their location. We could not identify any significant advantage in using statistical tools such as linear regression or Bayesian networks or ad-hoc methods [15].

Uncovering Causal Relationships between Software Metrics and Bugs Bug prediction is an important challenge for software engineering research that consists in looking for possible early indicators of the presence of bugs in a software. However, despite the relevance of the issue, most experiments designed to evaluate bug prediction only investigate whether there is a linear relation between the predictor and the presence of bugs. However, it is well known that standard regression models can not filter out spurious relations. We describe an experiment to discover more robust evidences towards causality between software metrics (as predictors) and the occurrence of bugs. For this purpose, we have relied on Granger Causality Test to evaluate whether past changes in a given time series are useful to forecast changes in another series. As its name suggests, Granger Test is a better indication of causality between two variables. We present and discuss the results of experiments on four real world systems evaluated over a time frame of almost four years. Particularly, we have been able to discover in the history of metrics the causes - in the terms of the Granger Test - for 64% to 93% of the defects reported for the systems considered in our experiment [18].

BugMaps: A Tool for the Visual Exploration and Analysis of Bugs To harness the complexity of big legacy software, software engineering tools need more and more information on these systems. This information may come from analysis of the source code, study of execution traces, computing of metrics, etc. One source of information received less attention than source code: the bugs on the system. Little is known about the evolutionary behavior, lifetime, distribution, and stability of bugs. We propose to consider bugs as first class entities and a useful source of information that can answer such topics. Such analysis is inherently complex, because bugs are intangible, invisible, and difficult to be traced. Therefore, our tool extracts information about bugs from bug tracking systems, link this information to other software artifacts, and explore interactive visualizations of bugs that we call bug maps [19].

A Catalog of Patterns for Concept Lattice Interpretation in Software Reengineering Formal Concept Analysis (FCA) provides an important approach in software reengineering for software understanding, design anomalies detection and correction. However, FCA-based approaches have two problems: (i) they produce lattices that must be interpreted by the user according to his/her understanding of the technique and different elements of the graph; and, (ii) the lattice can rapidly become so big that one is overwhelmed by the mass of information and possibilities. We make a catalog of important patterns in concept lattices, which can allow automating the task of lattice interpretation. The approach helps the reengineer to concentrate on the task of reengineering rather than understanding a complex lattice. We provide interpretation of these patterns in a generalized manner and illustrate them on various contexts constructed from program information of different open-source systems. We also present a tool that allows automated extraction of the patterns from concept lattices [16].

5.4. Reflective

Participants: Marcus Denker [Correspondant], Stéphane Ducasse.

DynamicSchema: a lightweight persistency framework for context-oriented data management While context-oriented programming technology so far has focused mostly on behavioral adaptation, context-oriented data management has received much less attention. We make a case for the problem of context-oriented data management, using a concrete example of a mobile application. We illustrate some of the issues involved and propose a lightweight persistency framework, called DynamicSchema, that resolves some of these issues. The solution consists in a flexible reification of the database schema, as a convenient dynamic data structure that can be adapted at execution time, according to sensed context changes. Implementing our mobile application using this framework enabled us to reduce the complexity of the domain modeling layer, to facilitate the production of code with low memory footprint, and to simplify the implementation of certain scenarios related to context-dependent security concerns [17].

ROMA Team

5. New Results

5.1. Unified model for assessing checkpointing protocols at extreme-scale

In this work [38], we defined a unified model for several well-known checkpoint/restart protocols. The proposed model is generic enough to encompass both extremes of the checkpoint/restart space, from coordinated approaches to a variety of uncoordinated checkpoint strategies (with message logging). We identified a set of crucial parameters, instantiated them and compared the expected efficiency of the fault tolerant protocols, for a given application/platform pair. We then proposed a detailed analysis of several scenarios, including some of the most powerful currently available HPC platforms, as well as anticipated Exascale designs. The results of this analytical comparison are corroborated by a comprehensive set of simulations. Altogether, they outlined comparative behaviors of checkpoint strategies at very large scale, thereby providing insight that is hardly accessible to direct experimentation.

5.2. Impact of fault prediction on checkpointing strategies

We dealt [34] with the impact of fault prediction techniques on checkpointing strategies. We extended the classical analysis of Young and Daly in the presence of a fault prediction system, which is characterized by its recall and its precision, and which provides either exact or window-based time predictions. We succeeded in deriving the optimal value of the checkpointing period (thereby minimizing the waste of resource usage due to checkpoint overhead) in all scenarios. These results allow to analytically assess the key parameters that impact the performance of fault predictors at very large scale. In addition, the results of this analytical evaluation were nicely corroborated by a comprehensive set of simulations, thereby demonstrating the validity of the model and the accuracy of the results.

5.3. Combining process replication and checkpointing for resilience on exascale systems

Processor failures in post-petascale settings are common occurrences. The traditional fault-tolerance solution, checkpoint-rollback, severely limits parallel efficiency. One solution is to replicate application processes so that a processor failure does not necessarily imply an application failure. Process replication, combined with checkpoint-rollback, has been recently advocated by Ferreira et al. [52]. We first identified [41] an incorrect analogy made in their work between process replication and the birthday problem, and derived correct values for the Mean Number of Failures To Interruption and Mean Time To Interruption for exponentially distributed failures. We then extended these results to arbitrary failure distributions, including closed-form solutions for Weibull distributions. Finally, we evaluated process replication using both synthetic and real-world failure traces. Our main findings are: (i) replication is less beneficial than claimed by Ferreira et al.; (ii) although the choice of the checkpointing period can have a high impact on application execution in the no-replication case, with process replication this choice is no longer critical.

5.4. On the complexity of scheduling checkpoints for computational workflows

This work [22] dealt with the complexity of scheduling computational workflows in the presence of Exponential failures. When such a failure occurs, rollback and recovery is used so that the execution can resume from the last checkpointed state. The goal is to minimize the expected execution time, and we have to decide in which order to execute the tasks, and whether to checkpoint or not after the completion of each given task. We showed that this scheduling problem is strongly NP-complete, and proposed a (polynomial-time) dynamic programming algorithm for the case where the application graph is a linear chain. These results laid the theoretical foundations of the problem, and constituted a prerequisite before discussing scheduling strategies for arbitrary DAGS of moldable tasks subject to general failure distributions.

5.5. Scheduling tree-shaped task graphs to minimize memory and makespan

We [44] investigated the execution of tree-shaped task graphs using multiple processors. Each edge of such a tree represents a large IO file. A task can only be executed if all input and output files fit into memory, and a file can only be removed from memory after it has been consumed. Such trees arise, for instance, in the multifrontal method of sparse matrix factorization. The maximum amount of memory needed depends on the execution order of the tasks. With one processor the objective of the tree traversal is to minimize the required memory. This problem was well studied and optimal polynomial algorithms were proposed. We extended the problem by considering multiple processors, which is of obvious interest in the application area of matrix factorization. With the multiple processors comes the additional objective to minimize the time needed to traverse the tree, i.e., to minimize the makespan. Not surprisingly, this problem proves to be much harder than the sequential one. We studied the computational complexity of this problem and provided an inapproximability result even for unit weight trees. We proposed several heuristics, each with a different optimization focus, and we analyzed them in an extensive experimental evaluation using realistic trees.

5.6. Memory allocation for different classes of DAGs

In this work, we studied the complexity of traversing workflows whose tasks require large I/O files. Such workflows arise in many scientific fields, such as image processing, genomics or geophysical simulations. They usually exhibit some regularity, and most of them can be modeled as Series-Parallel Graph. We target a classical two-level memory system, where the main memory is faster but smaller than the secondary memory. A task in the workflow can be processed if all its predecessors have been processed, and if its input and output files fit in the currently available main memory. The amount of available memory at a given time depends upon the ordering in which the tasks are executed. We focus on the problem of minimizing the amount of main memory needed to process the whole DAG.

We first concentrate on the parallel composition of task chains, or fork-join graphs. We adapt an algorithm designed for trees by Liu [54]. We prove that an optimal schedule for fork-join can be split in two optimal tree schedules, which are obtained using Liu's algorithm. We then move to Series-Parallel graphs and propose a recursive adaptation of the previous algorithm, which consists in serializing every parallel compositions, starting from the innermost, using the fork-join algorithm. Simulations show that this algorithm always reach the optimal performance, and we provide a sketch of the optimality proof. We also study compositions of complete bipartite graphs, which are another important class of DAGs arising in scientific workflows. We propose an optimal algorithm for a class of compositions which we name tower of complete bipartite graphs.

5.7. Scheduling non-linear divisible loads

Divisible Load Theory (DLT) has received a lot of attention in the past decade. A divisible load is a perfect parallel task, that can be split arbitrarily and executed in parallel on a set of possibly heterogeneous resources. The success of DLT is strongly related to the existence of many optimal resource allocation and scheduling algorithms, what strongly differs from general scheduling theory. Moreover, recently, close relationships have been underlined between DLT, that provides a fruitful theoretical framework for scheduling jobs on heterogeneous platforms, and MapReduce, that provides a simple and efficient programming framework to deploy applications on large scale distributed platforms.

The success of both have suggested to extend their framework to non-linear complexity tasks. We show [35] that both DLT and MapReduce are better suited to workloads with linear complexity. In particular, we prove that divisible load theory cannot directly be applied to quadratic workloads, such as it has been proposed recently. We precisely state the limits for classical DLT studies and we review and propose solutions based on a careful preparation of the dataset and clever data partitioning algorithms. In particular, through simulations, we show the possible impact of this approach on the volume of communications generated by MapReduce, in the context of Matrix Multiplication and Outer Product algorithms.

5.8. Energy-aware scheduling under reliability and makespan constraints

We consider [13] a task graph mapped on a set of homogeneous processors. We aim at minimizing the energy consumption while enforcing two constraints: a prescribed bound on the execution time (or makespan), and a reliability threshold. Dynamic voltage and frequency scaling (DVFS) is an approach frequently used to reduce the energy consumption of a schedule, but slowing down the execution of a task to save energy is decreasing the reliability of the execution.

In this work, to improve the reliability of a schedule while reducing the energy consumption, we allow for the re-execution of some tasks. We assess the complexity of the tri-criteria scheduling problem (makespan, reliability, energy) of deciding which task to re-execute, and at which speed each execution of a task should be done, with two different speed models: either processors can have arbitrary speeds (continuous model), or a processor can run at a finite number of different speeds and change its speed during a computation (VDD model). We propose several novel tri-criteria scheduling heuristics under the continuous speed model, and we evaluate them through a set of simulations. The two best heuristics turn out to be very efficient and complementary.

5.9. Approximation algorithms for energy, reliability and makespan optimization problems

We consider [32] the problem of scheduling an application on a parallel computational platform. The application is a particular task graph, either a linear chain of tasks, or a set of independent tasks. The platform is made of identical processors, whose speed can be dynamically modified. It is also subject to failures: if a processor is slowed down to decrease the energy consumption, it has a higher chance to fail. Therefore, the scheduling problem requires to re-execute or replicate tasks (i.e., execute twice a same task, either on the same processor, or on two distinct processors), in order to increase the reliability. It is a tri-criteria problem: the goal is to minimize the energy consumption, while enforcing a bound on the total execution time (the makespan), and a constraint on the reliability of each task.

Our main contribution is to propose approximation algorithms for these particular classes of task graphs. For linear chains, we design a fully polynomial time approximation scheme. However, we show that there exists no constant factor approximation algorithm for independent tasks, unless P=NP, and we are able in this case to propose an approximation algorithm with a relaxation on the makespan constraint.

5.10. Optimal algorithms and approximation algorithms for replica placement with distance constraints in tree networks

We study [16] the problem of replica placement in tree networks subject to server capacity and distance constraints. The client requests are known beforehand, while the number and location of the servers are to be determined. The Single policy enforces that all requests of a client are served by a single server in the tree, while in the Multiple policy, the requests of a given client can be processed by multiple servers, thus distributing the processing of requests over the platform. For the Single policy, we prove that all instances of the problem are NP-hard, and we propose approximation algorithms. The problem with the Multiple policy was known to be NP-hard with distance constraints, but we provide a polynomial time optimal algorithm to solve the problem in the particular case of binary trees when no request exceeds the server capacity.

5.11. Throughput optimization for pipeline workflow scheduling with setup times

We tackle [15] pipeline workflow applications that are executed on a distributed platform with setup times. In such applications, several computation stages are interconnected as a linear application graph, and each stage holds a buffer of limited size where intermediate results are stored and a processor setup time occurs when passing from one stage to another. The considered stage/processor mapping strategy is based on interval mappings, where an interval of consecutive stages is performed by the same processor and the objective is the throughput optimization. Typical examples for this kind of applications are streaming applications such as audio and video coding or decoding, image processing using co-processing devices as FPGA. Even when neglecting setup times, the problem is NP-hard on heterogeneous platforms and we therefore restrict to homogeneous resources. We provide an optimal algorithm for constellations with identical buffer capacities. When buffer sizes are not fixed, we deal with the problem of allocating the buffers in shared memory and present a $b/(b+1)$-approximation algorithm.

5.12. Semi-matching algorithms for scheduling parallel tasks under resource constraints

We study [37] the problem of minimum makespan scheduling when tasks are restricted to subsets of the processors (resource constraints), and require either one or multiple distinct processors to be executed (parallel tasks). This problem is related to the minimum makespan scheduling problem on unrelated machines, as well as to the concurrent job shop problem, and it amounts to finding a semi-matching in bipartite graphs or hypergraphs. While the problem was known to be NP-complete for bipartite graphs, but solvable in polynomial time for unweighted graphs (i.e., unit tasks), we prove that the problem is NP-complete for hypergraphs even in the unweighted case. We design several greedy algorithms of low complexity to solve two versions of the problem, and assess their performance through a set of exhaustive simulations. Even though there is no approximation guarantee on these linear algorithms, they return solutions close to the optimal (or a known lower bound) in average.

5.13. A Symmetry preserving algorithm for matrix scaling

We present an iterative algorithm which asymptotically scales the ∞-norm of each row and each column of a matrix to one. This scaling algorithm preserves symmetry of the original matrix and shows fast linear convergence with an asymptotic rate of $1/2$. We discuss extensions of the algorithm to the one-norm, and by inference to other norms. For the 1-norm case, we show again that convergence is linear, with the rate dependent on the spectrum of the scaled matrix. We demonstrate experimentally that the scaling algorithm improves the conditioning of the matrix and that it helps direct solvers by reducing the need for pivoting. In particular, for symmetric matrices the theoretical and experimental results highlight the potential of the proposed algorithm over existing alternatives. This work resulted in an improved version [43] of an earlier technical report [55].

5.14. On shared-memory parallelization of a sparse matrix scaling algorithm

We discuss [25] efficient shared memory parallelization of sparse matrix computations whose main traits resemble to those of the sparse matrix-vector multiply operation. Such computations are difficult to parallelize because of the relatively small computational granularity characterized by small number of operations per each data access. Our main application is a sparse matrix scaling algorithm which is more memory bound than the sparse matrix vector multiplication operation. We take the application and parallelize it using the standard OpenMP programming principles. Apart from the common race condition avoiding constructs, we do not reorganize the algorithm. Rather, we identify associated performance metrics and describe models to optimize them. By using these models, we implement parallel matrix scaling algorithms for two well-known sparse matrix storage formats. Experimental results show that simple parallelization attempts which leave data/work partitioning to the runtime scheduler can suffer from the overhead of avoiding race conditions especially when the number of threads increases. The proposed algorithms perform better than these algorithms by optimizing the identified performance metrics and reducing the overhead.

5.15. Investigations on push-relabel based algorithms for the maximum transversal problem

In a technical report [42], we investigate the push-relabel algorithm for solving the problem of finding a maximum cardinality matching in a bipartite graph in the context of the maximum transversal problem. We describe in detail an optimized yet easy-to-implement version of the algorithm and fine-tune its parameters. We also introduce new performance-enhancing techniques. On a wide range of real-world instances, we compare the push-relabel algorithm with state-of-the-art augmenting path-based algorithms and the recently proposed pseudoflow approach. We conclude that a carefully tuned push-relabel algorithm is competitive with all known augmenting path-based algorithms, and superior to the pseudoflow-based ones. We finalized this work by reporting the most important results in a journal article [9].

5.16. On optimal and balanced sparse matrix partitioning problems

We investigate [20] one dimensional partitioning of sparse matrices under a given ordering of the rows/columns. The partitioning constraint is to have load balance across processors when different parts are assigned to different processors. The load is defined as the number of rows, or columns, or the nonzeros assigned to a processor. The partitioning objective is to optimize different functions, including the well-known total communication volume arising in a distributed memory implementation of parallel sparse matrix-vector multiplication operations. The difference between our problem in this work and the general sparse matrix partitioning problem is that the parts should correspond to disjoint intervals of the given order. Whereas the partitioning problem without the interval constraint corresponds to the NP-complete hypergraph partitioning problem, the restricted problem corresponds to a polynomial-time solvable variant of the hypergraph partitioning problem. We adapt an existing dynamic programming algorithm designed for graphs to solve two related partitioning problems in graphs. We then propose graph models for a given hypergraph and a partitioning objective function so that the standard cutsize definition in the graph model exactly corresponds to the hypergraph partitioning objective function. In extensive experiments, we show that our proposed algorithm is helpful in practice. It even demonstrates performance superior to the standard hypergraph partitioners when the number of parts is high.

5.17. Constructing elimination trees for sparse unsymmetric matrices

The elimination tree model for sparse unsymmetric matrices and an algorithm for constructing it have been recently proposed [50], [51]. The construction algorithm has a worst-case time complexity of $\Theta(mn)$ for an $n \times n$ unsymmetric matrix having m off-diagonal nonzeros. We proposed [53] another algorithm that has a worst-case time complexity of $\mathcal{O}(m \log n)$. During this reporting period, we compared the two algorithms experimentally and showed that both algorithms are efficient in general. The known algorithm [51] is faster in many practical cases, yet there are instances in which there is a significant difference between the running time of the two algorithms in favor of the proposed one.

5.18. Introduction of shared memory parallelism in a distributed-memory sparse multifrontal solver

We study the adaptation of a parallel distributed-memory solver, MUMPS, into a shared-memory code, targetting multicore architectures. An advantage of adapting the code rather than starting with a new design is to fully benefit from its numerical kernels and functionalities. We show how one can take advantage of OpenMP directives and of existing libraries optimized for shared-memory environments, in our case BLAS libraries [48]. We have also started to study approaches that take advantage of the specificities of NUMA architectures.

5.19. Improving multifrontal methods by means of low-Rank representations

Matrices coming from elliptic PDEs have been shown to have a low-rank property. Although the dense internal datastructures involved in a multifrontal method, the so-called frontal matrices or fronts, are full-rank, their off-diagonal blocks can then be approximated by low-rank products. We have studied a low-rank format called Block Low Rank and explained how it can be used to reduce the memory footprint and complexity of both the factorization and solve phases, depending on the way variables are grouped. The proposed approach can be used either to accelerate the factorization and solution phases or to build a preconditioner [47]. We have started the development of a version of MUMPS that exploits such properties. This work is in collaboration with EDF (contract funding for the Ph.D. thesis of C. Weisbecker at INPT) and C. Ashcraft (LSTC).

5.20. Parallel computation of inverse entries of a sparse matrix

We have worked on the parallel computation of several entries [31] of the inverse of a large sparse matrix. We assume that the matrix has already been factorized by a direct method and that the factors are distributed. Entries are efficiently computed by exploiting sparsity of the right-hand sides and the solution vectors in the triangular solution phase. We demonstrate that in this setting, parallelism and computational efficiency are two contrasting objectives. We develop an efficient approach and show its efficacy by runs using the MUMPS code that implements a parallel multifrontal method.

5.21. Robust memory-aware mappings for parallel multifrontal factorization

We have studied the memory scalability of the parallel multifrontal factorization of sparse matrices. In particular, we are interested in controlling the active memory specific to the multifrontal factorization. We illustrate why commonly used mapping strategies (e.g. proportional mapping) cannot achieve a high memory efficiency. We propose a class of "memory-aware" algorithms that aim at maximizing performance under given memory constraints, and explain why they provide reliable memory estimates, thus a more robust solver. We study these issues in the context of the MUMPS solver, in which new experimental static scheduling strategies have been implemented and experimented on large matrices [46].

RUNTIME Project-Team

6. New Results

6.1. Mastering Heterogeneous Platforms

Participants: Cedric Augonnet, Olivier Aumage, Nicolas Collin, Ludovic Courtès, Nathalie Furmento, Sylvain Henry, Andra Hugo, Raymond Namyst, Cyril Roelandt, Corentin Rossignon, Ludovic Stordeur, Samuel Thibault, Pierre-André Wacrenier.

- We continued our work on extending STARPU to master exploitation of Heterogeneous Platforms.
- We have released version 1.0.0 of STARPU, now really considered a stable project that a lot of collaborators can base their work on.
- We have extended our lightweight DSM over MPI to support caching data [17], which dramatically reduces data transfers for classical applications.
- We have extended the STARPU scheduler to let the application provide several implementations of a function for the same architecture, implementation choice being performed by the scheduler according to actually measured performance, energy consumption, etc.
- We have collaborated with Computer Graphics research team in the MediaGPU project to make it possible to directly graphically render results from STARPU computations.
- Work has been initiated to integrate STARPU and SIMGRID for the SONGS project, which will allow to simulate application execution on heterogeneous architectures, and thus easily experiment with scheduling strategies.
- We have extended STARPU with a protocol that permits to make it run with a master-slave model, which allowed to easily port it to the Intel SCC and Intel Xeon Phi processors, and will allow an easy load balancing support over MPI.
- We have extended STARPU to allow multiple parallel codes to run concurrently with minimal interference. Such parallel codes run within *scheduling contexts* that provide confined execution environments which can be used to partition computing resources. Scheduling contexts can be dynamically resized to optimize the allocation of computing resources among concurrently running libraries. We introduced a *hypervisor* that automatically expands or shrinks contexts using feedback from the runtime system (e.g. resource utilization).

  We demonstrated the relevance of our approach using benchmarks invoking multiple high performance linear algebra kernels simultaneously on top of heterogeneous multicore machines. We showed that our mechanism can dramatically improve the overall application run time (-34%), most notably by reducing the average cache miss ratio (-50%).
- We have improved [15] the OPENCL implementation on top of StarPU (SOCL) to allow applications to use STARPU's scheduling contexts through OPENCL's contexts and to explicitly schedule some kernels to enhance performance. Moreover, SOCL fully supports the OPENCL ICD extension and can now be dynamically selected amongst other available platforms which makes it easier to use.
- We have continued collaborations on applications on top of STARPU with the University of Mons [14], the University of Vienna [20], the University of Linköping, the University of Tsukuba, TOTAL, the CEA INAC in Grenoble and the BRGM French public institution in Earth science applications.
- In a joint work with French SME company CAPS entreprise, as part of the ANR ProHMPT project, we have demonstrated a proof of concept framework enabling three kinds of pieces of applicative code — a native StarPU code, a Magma/StarPU code and a HMPP/StarPU code annotated with HMPP's directives — to integrate and cooperate together on a computation as a single coherent application.

- As part of the HPC-GA project, we initiated a preliminary study with University of Rio Grande do Sul (UFRGS), Brazil, to cooperate on the modeling of common computing kernel tasks and potentially making use of kernel models designed at UFRGS within the StarPU's task cost evaluation framework.
- As part of the partnership with Total, and in relationship with StarPU's task scheduling work, we have explored solutions to semi-automatically adapt the grain of elementary tasks to the available computing resources.

6.2. High-Performance Intra-node Collective Operations

Participant: Brice Goglin.

- KNEM is known to improve the performance of point-to-point intra-node MPI communication significantly [13].
- We designed an extended RMA interface in KNEM that suits the needs of point-to-point, collective and RMA operations.
- We showed that the native use of KNEM in MPI collective implementations enabled further optimization by combining the knowledge of collective algorithms with the mastering of KNEM region management and copies.
- This work was initiated in the context of our collaboration with the MPICH2 team and is now also pursued within the OPEN MPI project in collaboration with the University of Tennessee in Knoxville.

6.3. Process Placement and Topology-Aware Computing

Participants: Emmanuel Jeannot, Guillaume Mercier, François Tessier.

- TREEMATCH's limitations have been addressed. In particular, it is now able to handle unbalanced physical topologies.
- TREEMATCH has been compared to various competitors. We carried out various experiments that showed that TREEMATCH outperforms other solutions based on graph partitioning or graph embedding. These experiments also showed the limitations of some existing solutions (Scotch for instance).
- TREEMATCH has been integrated into several major parallel programming environments. It is implemented as a load-balancer in Charm++ (François TESSIER made severals at UI Urbana Champaign) and is used to enhance topology management routines in Open MPI and MPICH2. It is indeed employed to allow rank reordering in functions such as `MPI_Dist_graph_create` for instance. This work started with a visit at UTK by Guillaume MERCIER.
- We set-up several collaborations: besides the collaboration with the Open MPI group, we also work with the CERFACS in order to speed-up existing CFD parallel applications developed by this group.

6.4. Thread placement and memory allocation on NUMA machines

Participant: Emmanuel Jeannot.

We have worked on optimizing the tiled Cholesky factorization on NUMA machine. We have designed a new symbolic technnique for allocating task and tiles at the same time called SMA (Symbolic Mapping and Allocation). SMA provide an optimal allocation in terms of point-to-point communication for the Cholesky factorization. We have studied some performance issues regarding the way threads are grouped and tiles are allocated in the memory. We have shown how to optimize thread placement and data placement in order to achieve performance gain up to 50% compared to state-of-the-art libraries such as Plasma or MKL. This work has been published in PAAP 2012 [25].

6.5. Scheduling for System On Chip

Participants: Paul-Antoine Arras, Emmanuel Jeannot, Samuel Thibault.

Today's embedded applications are increasingly demanding in terms of computational power, especially in real-time digital signal processing (DSP) where tight timing requirements are to be fulfilled. More specifically, when it comes to video decoding (e.g. H.264/AVC and HEVC) not only has it been almost impossible for some time to run such codecs on a stand-alone embedded processor, but it now also becomes quite impractical to execute them on homogeneous multicore platforms. In this context, STMicroelectronics is developing a scalable heterogeneous system-on-chip template called STHORM and aimed at meeting the latest codecs' requirements.

This year, we focused on the memory constraints embedded systems are subject to. As video coding is rather demanding in terms of storage capacity, we have proposed a method aimed at introducing the notion of memory into a class of widespread scheduling heuristics that exhibit both good performance and low complexity. Thanks to this technique, we achieved speedups over 20%.

The next step is to formalize an execution model on top of which a runtime software will be built. This implies specifying both the application requirements and modeling precisely the target platform, namely STHORM.

6.6. High-Performance Point-to-Point Communications

Participants: Alexandre Denis, Sébastien Barascou, Raymond Namyst.

- NEWMADELEINE is our communication library designed for high performance networks in clusters. We have worked on optimizations on low-level protocols so as to improve point-to-point performance.
- We have proposed a communication protocol [21] for InfiniBand that amortizes the cost of checksums as used by fault-tolerant MPI implementations. We have modeled the behavior of the network and proposed auto-tuning mechanisms to adapt the protocol to the hardware properties.
- This work was initiated in the context of the FP3C collaboration with the University of Tokyo.

SARDES Project-Team

5. New Results

5.1. Languages and Foundations: Process algebra

Participants: Damien Pous, Jean-Bernard Stefani, Barbara Petit.

The goal of this work is to study process algebraic foundations for component-based distributed programming. Most of this work takes place in the context of the ANR PiCoq and Rever projects.

To develop composable abstractions for programming dependable systems, we investigate concurrent reversible models of computation, where arbitrary executions can be reversed, step by step, in a causally consistent way. This year we have continued the study of primitives for controlling reversibility in a higher-order variant of the π-calculus. We have shown that the combination of a basic notion of message alternative coupled with a rollback primitive that respects causal consistency provides enough expresive power to encode various rollback policies. We have also started to study the expresive power of these primitives with respect to transactional constructs. In particular, we have shown that our primitives allow for a faithful encoding of a notion of communicating transaction proposed by Hennessy et al, while avoiding spurious rollbacks which mar Hennessy's approach. This work has been submitted for publication. A digest of our main ideas on controlling reversibility has appeared in [25].

We have also started a study on the cost of making a concurrent programming language reversible. More specifically, we have started from an abstract machine for a fragment of the Oz programming language and made it reversible. We have shown that the overhead of the reversible machine with respect to the original one in terms of space is at most linear in the number of execution steps, and that this bound is tight since some programs cannot be made reversible without storing a commensurate amount of information. This work has been published in [26].

5.2. Control for adaptive systems: Discrete control for adaptive and reconfigurable systems

Participants: Eric Rutten, Noël De Palma, Olivier Gruber, Fabienne Boyer, Xin An, Soguy Mak-Kare Gueye.

The goal of this work is to apply control techniques based on the behavioral model of reactive automata and the algorithmic techniques of discrete controller synthesis. We adopt the synchronous approach to reactive systems, and use an associated effective controller synthesis tool, Sigali, developed at Inria Rennes. We are exploring several target application domains, where we expect to find commonalities in the control problems, and variations in the definitions of configurations, and in the criteria motivating adaptation.

This year, we have started investigating the application of discrete controller synthesis to various problems in computer systems management and administration. The increasing complexity of computer systems has led to the automation of administration functions, in the form of autonomic managers. One important aspect requiring such management is the issue of energy consumption of computing systems, in the perspective of green computing. As these managers address each a specific aspect, there is a need for using several managers to cover all the domains of administration. However, coordinating them is necessary for proper and effective global administration. Such coordination is a problem of synchronization and logical control of administration operations that can be applied by autonomous managers on the managed system at a given time in response to events observed on the state of this system. We therefore propose to investigate the use of reactive models with events and states, and discrete control techniques to solve this problem. In [20], [21], [31], [30], we illustrate this approach by integrating a controller obtained by synchronous programming, based on Discrete Controller Synthesis, in an autonomic system administration infrastructure. The role of this controller is to orchestrate the execution of reconfiguration operations of all administration policies to satisfy properties of logical consistency. We have applied this approach to coordinate energy-aware managers for self-optimization, self-regulation of processor frequency and self-repair.

5.3. System support: System support for multicore machines

Participants: Vivien Quéma, Renaud Lachaize, Baptiste Lepers.

Multicore machines with Non-Uniform Memory Accesses (NUMA) are becoming commodity platforms. Efficiently exploiting their resources remains an open research problem. This line of work investigates system support to tackle various issues related to efficient resource management and programming support.

One of the key concerns in efficiently exploiting multicore NUMA architectures is to limit as much as possible the number of remote memory accesses (i.e., main memory accesses performed from a core to a memory bank that is not directly attached to it). However, in many cases, existing profilers do not provide enough information to help programmers achieve this goal. We have developed MemProf [24], the first profiler that allows programmers to choose and implement efficient application-level optimizations for NUMA systems. MemProf achieves this goal by allowing programmers to (i) precisely understand which memory objects are accessed remotely in memory, and (ii) building temporal flows of interactions between threads and objects. We evaluated MemProf using four applications (FaceRec, Streamcluster, Psearchy, and Apache) on three different machines. In each case, we showed how MemProf helped us choose and implement efficient optimizations, unlike existing profilers. These optimizations provide significant performance gains on the studied applications (up to 161%), while requiring very lightweight modifications (10 lines of code or less).

State-machine replication is a well-known fault-tolerance technique. Unfortunately existing state-machine replication schemes do not scale well on multicore machines. In collaboration with U. Texas at Austin (L. Alvisi), we have developed a new state-machine replication scheme [23], that departs from the standard agree-execute architecture of existing schemes, in favor of a more optimistic, and less deterministic, execute-verify replication scheme, which yields much better scalability. We have evaluated Eve's throughput gain compared with traditional sequential execution approaches, as well as Eve's overheads compared to unreplicated multithreaded execution and to alternative replication approaches.

5.4. System support: Performance and dependability benchmaking

Participants: Amit Sangroya, Damian Serrano-Garcia, Sara Bouchenak [correspondant].

MapReduce is a popular programming model for distributed data processing. Extensive research has been conducted on the reliability of MapReduce, ranging from adaptive and on-demand fault-tolerance to new fault-tolerance models. However, realistic benchmarks are still missing to analyze and compare the effectiveness of these proposals. To date, most MapReduce fault-tolerance solutions have been evaluated using micro benchmarks in an ad-hoc and overly simplified setting, which may not be representative of real-world applications. To remedy this situation, we have developed MRBS, a comprehensive benchmark suite for evaluating the dependability of MapReduce systems. MRBS includes five benchmarks covering several application domains and a wide range of execution scenarios such as data-intensive vs. compute-intensive applications, or batch applications vs. online interactive applications. MRBS allows to inject various types of faults at different rates. It also considers different application workloads and data loads, and produces extensive reliability, availability and performance statistics. We have shown the use of MRBS with Hadoop clusters running on Amazon EC2, and on a private cloud [29], [28].

SCORE Team

6. New Results

6.1. Collaborative Data Management

6.1.1. A Framework to Design Conflict-Free Replicated Data Types

Participants: Mehdi Ahmed-Nacer, Stéphane Martin, Pascal Urso.

Design new eventually consistent data types is difficult and error-prone as demonstrated by the numerous proposed approaches that fail to resolve conflicts for simple plain text document. Moreover, more the data type is complex, more conflicts types must be resolved. We have presented a layered approach to design new eventually consistent data types [21], [15]. This approach decouples eventual consistency management from data type constraints satisfaction. We compose one or several existing replicated data types which ensure eventual consistency, and adaptation layers to obtain a new eventually consistent data type. Each layer or replicated data type can be freely substituted by one providing the same interface. We have demonstrated that our approach is implementable and obtains acceptable performances. Our experiments and implementation are publicly available and re-playable (https://github.com/score-team/replication-benchmarker).

6.1.2. Enhancing Rich Content Wikis with Real-Time Collaboration

Participants: Luc André, Claudia-Lavinia Ignat, Gérald Oster.

Wikis are one of the most important tools of Web 2.0 allowing users to easily edit shared data. WYSIWYG editors for wiki pages avoid the impediments of learning wiki syntax. However, wikis offer poor support for merging concurrent contributions on the same pages. Users have to manually merge concurrent changes and there is no support for an automatic merging. As real-time collaborative editing reduces the number of conflicts as the time frame for concurrent work is very short, we proposed extending wiki systems with real-time collaboration [23]. We propose an automatic merging solution adapted for rich content wikis. Our solution is integrated as an extension of XWiki system (http://extensions.xwiki.org/xwiki/bin/view/Extension/RealTime+Wiki+Editor).

6.1.3. Rapid and Round-free Multi-pair Asynchronous Push-Pull Aggregation

Participants: Claudia-Lavinia Ignat, Hyun-Gul Roh.

In the context of STREAMS project we investigated gossip-based dissemination mechanisms in peer-to-peer real-time collaboration adapted for consistency maintenance algorithms based on CRDT (Commutative Replicated Data Types). These dissemination mechanisms need to compute the size of the network and therefore a suitable rapid protocol that aggregates data over network is essential. Iterative aggregation protocols, especially push-pull style aggregations, generally need prior configurations to synchronize rounds over all nodes, and messages should be exchanged in a synchronous/blocking way in order to ensure accurate estimates in push-pull or push-sum protocols. We proposed a multi-pair asynchronous push-pull aggregation (MAPPA) [22], which frees the push-pull aggregations from the synchronization constraints, and therefore accelerates the aggregation speed. MAPPA is resilient to network churns, and thus suitable for dynamic peer-to-peer networks.

6.1.4. Trustworthy contract based collaboration

Participants: Claudia-Lavinia Ignat, Hien Thi Thu Truong.

Availability of trustworthy environments is one of the main conditions that would lead to a greater acceptance and reliance on collaborative systems. In the context of large scale multi-synchronous collaboration where users work in parallel on different streams of activities a ”hard” security that would forbid many actions is unusable. We adopt instead a “soft” security where rather than adopting an a priori strict enforcement of security rules, access is given first to data without control but with restrictions that are verified a posteriori. We proposed a contract-based collaboration model [2], [4] where we establish and adjust trust in users based on detective enforcement of basic usage control requirements. Usage control requirements are specified as contracts. Contracts are specified by data owners when they share data in accordance with user trust levels. Observation of adherence to or violation of contracts is used to adjust trust levels. Our contract-based collaboration model allows the specification of contracts, merging of data and contracts and resolution of conflicting contracts. A trust metric for computing user trust levels was proposed based on auditing user compliance to the given contracts.

Multi-synchronous collaboration maintains multiple, simultaneous streams of activity which continually diverge and converge. These streams of activity are represented by means of logs of operations, i.e. user modifications. A malicious user might tamper his log of operations. At the moment of synchronization with other streams, the tampered log might generate wrong results. A trustworthy collaboration environment should detect if logs were tampered. We proposed a mechanism for establishment of trusted logs relying on hash-chain based authenticators [17], [18], [2]. Our solution ensures the authenticity, the integrity of logs, and the user accountability. We proposed algorithms to construct authenticators and verify logs. We proved their correctness and provided theoretical and practical evaluations.

6.1.5. Distributed activity management in crisis situation

Participants: François Charoy, Joern Franke.

Crisis management has been a very fruitful domain to investigate new approaches for high value, human driven activity coordination in a multi organisational setting. Our work benefits from a large amount of use cases and detailed accounts of previous dramatic events to analyse requirements and confront our proposals. This paper present the final part of this work on the problem of replication of activities between several workspaces [3]. We are now looking for new vehicles to continue this research at an international level.

6.2. Data Centered Service Oriented Computing

6.2.1. Business process distribution on a SaaS architecture

Participants: Walid Fdhila, Claude Godart, Elio Goettelmann, Samir Youcef.

The objective of this work is to support the deployment of a business process as a set of distributed services provided partially or totally off-premises or even in the cloud. Direct applications in our target are:

- A methodological approach for choreographies elicitations and monitoring[12].
- An algorithm for optimized service providers selection (including cloud) [11], [9], [10].

In this objective, we have deployed two approaches. A first is based on heuristics (*greedy* algorithm to compute an initial solution, combined with a *tabu search*) for optimizing the selection of services assigned to activities in a decentralized composite service, both in terms of the overall QoS of the composite service and the communication overhead; in output, the initial business process model is translated in a set of interconnected business process fragments.

A second approach uses operational research techniques for optimizing a cloud selection taking into account two conflicting objectives, namely: the execution time (makespan) and the overall cost incurred using a set of resources. We proposed in [9] three complementary approaches to deal with the matching and scheduling scientific workflow tasks in Cloud computing environments. An extension of this first study was presented in [11], [10]. More precisely, we have extended the three proposed approaches to consider: (i) the business workflows and (ii) the concurrent access to resources by multiple instances of a given process. To achieve this goal, we proposed to use a predictive models in order to estimate the availability of the used resources. We are currently working on the business processes execution in Cloud computing context taking into account workflow patterns such as *sequence, switch, multi-choice, etc.* patterns. Moreover, we plan to extend the proposed work to take into account others criteria like carbon emission and energy cost.

6.2.2. Alignment between Business Process and Service Architecture

Participants: François Charoy, Karim Dahman, Claude Godart.

In the continuation of work done previously on change management during process execution, we are conducting work on the governance of change at the business level and on its implications at the architecture and infrastructure level of an information system. Last year was devoted to the definition of the transformation rules that allowed to go from a business model to an IT model, i.e. a transformation between model based on different paradigms. During this year, a great deal of effort has been done in order to extend our work on Business to IT alignment management. Our goal is still to maintain this alignment at the lowest possible cost when the business process are changing. Further than that we are trying to describe and validated an engineering method to help designer to maintain this alignment. Karim Dahman has defended is PhD on this matter in october 2012.

6.2.3. Monitoring and violations detections of choreographies or distributed compositions of services

Participants: Aymen Baouab, Ehtesham Zahoor, Olivier Perrin, Walid Fdhila, Claude Godart.

The dynamic nature of the cross-organizational business processes poses various challenges to their successful execution. Services choreographies or distributed compositions of services help to reduce such complexity by providing means for describing complex systems at a higher level. However, this does not necessarily guarantee that erroneous situations cannot occur due to inappropriately specified interactions. In [7], [6], we propose an approach for decentralized monitoring of cross-organizational choreographies, using a runtime event-based approach to deal with the problem of monitoring conformance of interaction sequences. Our approach allows for an automatic and optimized generation of rules. After parsing the choreography graph into a hierarchy of *canonical* blocks, tagging each event by its block ascendancy, an optimized set of monitoring queries is generated. We evaluate the concepts based on a scenario showing how much the number of queries can be significantly reduced. These results use our previous results about event-based framework DISC [33].

SOCRATE Team

6. New Results

6.1. Flexible Radio Node

6.1.1. Radio wave propagation

The MR-FDPF (Multi-Resolution Frequency Domain Partial Flow) method is proven to be a fast and efficient method to simulate radio wave propagation. It is a deterministic model which can provide an accurate radio coverage prediction. In reality, radio channels have the nature of randomness due to e.g. moving people or air flow. Thus they can not be rigorously simulated by a pure deterministic model. However, it is believed that some statistics can be extracted from deterministic models and these statistics can be very useful to describe radio channels in reality. In [20], large scale fading statistical characteristics are extracted based on the MR-FDPF method. They are validated by comparison to both the theoretical result and measurement. The match also demonstrates that MR-FDPF is capable of simulating large scale fading.

In [2] we study Realistic Prediction of Bit error rate (BER) and adaptive modulation and coding (AMC) for Indoor Wireless Transmissions. Bit error rate is an important parameter for evaluating the performance of wireless networks. In this letter, a realistic BER for indoor wireless transmissions is predicted. The prediction is based on a deterministic radio propagation model, the MR-FDPF model, which is capable of providing accurate fading statistics. The obtained BER map can be used in many cases, e.g., adaptive modulation and coding scheme or power allocation.

In [4], we propose a modification of the MR-FDPF method that allows simulating radio propagation channels in a frequency range. The performance of the proposed MR-FDPF implementation has been analyzed based on different realistic propagation scenarios. We also analyze the possibility of applying the multi-resolution frequency domain approach to the well-known transmission-line matrix method. The proposed multi-resolution frequency domain transmission-line matrix method provides a computationally efficient way of modeling radio wave propagation in three dimensional space at multiple frequencies.

In [3], we consider the performance of coded wireless communication systems experiencing non-frequency selective fading channels in shadowed environments. The quality of service (QoS) in a wireless network is dependent on the packet error outage (PEO). We address the problem of finding a tractable expression for the coded PEO over Nakagami-m channels with shadowing, considering multilevel modulations, various block, convolutional channel coding schemes and hard decision decoding. In order to obtain the coded PEO, an inversion of the coded packet error probability (PEP) w.r.t. the signal to noise ratio (SNR) is needed. To this end, we propose an invertible approximation for the coded PEP w.r.t. the uncoded bit error probability (BEP) in Nakagami-m fading channels which is accurate for all BEPs of interest. The BEP itself depends on the average SNR and we hence make use of previous results on the inversion of the uncoded BEP w.r.t. the SNR in Nakagamim fading channels, holding for M-PSK and M-QAM signals. We were thus able to obtain a reliable closed form expression for the coded PEO in flat fading and shadowing channels

6.1.2. Power consumption

In [24], we propose the use of an existing opensource network simulator, WSNet, to evaluate the interest of using multi-mode relays in terms of energy consumption. We show that the combination of MIMO and multi-mode provides a solution to reduce global energy consumption, but that conclusions are really scenario-dependent. Moreover, we explain how a multi-mode MIMO terminal can improve these results using adaptive strategies.

the energy consumption in wireless sensor networks is studied. In order to minimize the consumed power at the analog and RF part, an energy recovering system combined with a wake-up radio is proposed for discussion. The proposed architecture has three activity levels : zero consumption, low and high energy consumption. In order to quantify the gain in terms of power consumption, a power consumption model state of the art is proposed. in [7] all radio channel models which can be used for MIMO heterogeneous network with small cells are described.

6.1.3. MIMO

In [28], we study MIMO and next generation system. For the past decade or more MIMO systems have been the subject of very intensive research. However in the past few years, these techniques have begun to be implemented in practice. In particular they have appeared in the standards for next generation systems such as LTE, 3GPP-LTE Advanced and WiMAX, as well as the latest versions of Wifi. This chapter, extracted from the book edited by the Cost Action 2100: "Pervasive Mobile and Ambient Wireless Communications", brings together the MIMO systems used in next generation systems with other work on the implementation and simulation of these systems. It also describes advances in MIMO techniques in a number of areas. The first section is divided into two sub-sections dealing first with simulators and testbeds which are used in system-level simulators to evaluate overall system capacity, as discussed in later chapters of this book. Secondly the development of terminals for next generation MIMO systems is considered, especially considering the additional RF hardware required for MIMO. Section 7.2 then discusses especially precoding techniques used in many of the recent standards to implement MIMO. In particular precoding allows the implementation of closed loop or adaptive MIMO. In next generation systems there is also much increased attention on MU-MIMO and on multi-terminal MIMO in general, including so-called "network MIMO " approaches, which appear in LTE as Coordinated Multiple Point: this is covered in Sect. 7.3. Various advanced MIMO transmission and detection approaches are covered in Sects. 7.4 to 7.6, including some interesting work on MIMO techniques involving continuous phase modulation, giving advantages in terms of peak-to-average power ratio.

6.2. Agile Radio Resource Sharing

6.2.1. Wireless Multi-hop Networks

In [6], we study energy-delay tradeoff in wireless multihop networks with unreliable links. Energy efficiency and transmission delay are very important parameters for wireless multihop networks. Numerous works that study energy efficiency and delay are based on the assumption of reliable links. However, the unreliability of channels is inevitable in wireless multihop networks. In addition, most of works focus on self-organization protocol design while keeping non-protocol system parameters fixed. While, very few works reveal the relationship between the network performance and these physical parameters, in other words, the best networks performance could be obtained by the physical parameters. This paper investigates the tradeoff between the energy consumption and the latency of communications in a wireless multihop network using a realistic unreliable link model. It provides a closed-form expression of the lower bound of the energy–delay tradeoff and of energy efficiency for different channel models (additive white Gaussian noise, Rayleigh fast fading and Rayleigh block-fading) in a linear network. These analytical results are also verified in 2-dimensional Poisson networks using simulations. The closed-form expression provides a framework to evaluate the energy–delay performance and to optimize the parameters in physical layer, MAC layer and routing layer from the viewpoint of cross-layer design during the planning phase of a network.

6.2.2. Relay and Cooperative Communications

In [16], we aim at characterizing the gain induced by using relay channels in a linear network under both capacity constraint and realistic energy model. We express a general model based on a convex optimization problem. Then, we use numerical tools to obtain results on the outer and inner bounds of the capacity of the full and half duplex relay channel. We then extend this study with more complex networks based on relay channels, especially networks formed by a linear chain of nodes. We describe the Pareto optimal solutions of the minimization problem with respect to the consumed energy and latency in such a linear network. From the simple case of the linear multi-hop network, we study the gains when implementing a linear chain of relay channels and compare these results to the simpler multi-hop transmission.

In [15], we present preliminary results on achievables rates in half-duplex cooperative multiple access channels (CMAC).We show that the upper bound on the capacity of the half-duplex CMAC can be solved using convex optimization techniques. Under a Gaussian model, we study the maximal achievable rate by every node in the network. We propose a number of scenarios, encompassing existing and theoretical cooperation schemes. Using these hypotheses, we evaluate the performance of both a non-cooperative concurrent access and simple cooperative multi-hop or relaying schemes with respect to the upper bound. The performance is compared for the various scenarios, and we provide analyses of specific cases in order to illustrate how our framework may be used to answer targeted questions about the capacity of CMACs.

In [31], we aim at obtaining usable bounds on the performance of CMAC under a Gaussian model. We first show that the problem can be transformed into a convex optimization problem which makes it easily solvable using numerical tools. We propose, as a line of study, to consider the maximal achievable common rate by every node in the network. We then proceed to express closed-form bounds on the capacity region of the CMAC in that common rate scenario. We study simple cooperation schemes based on existing results in relay channels and compare them to other medium sharing approaches. In the end, we show that using the relay-channel based protocols can be efficient for some parameters, but gets less interesting in the Gaussian case if the source-destination links are good enough.

In [30], we study the optimal power allocations in CMACs, where we aim at maximizing the rate achievable by both sources simultaneously rather than the sum of achievable rates. Separating our study between the coherent and non-coherent case, we obtain closed-form expressions for the optimal power allocations w.r.t. the outer bounds of the capacity region, as well as decode-and-forward and non-cooperative inner bounds. We point out during our resolution that the general CMAC model behaves as a multiple access relay channel (MARC), where a "virtual" relay node is introduced to represent the cooperation between the sources. This equivalent model simplifies the original power allocation problem. We finally show that the general cut-set outer bound on the capacity region of the equivalent MARC matches exactly the tightest known outer bound on the capacity region of the original CMAC.

In [17] we address the distributed power adaptation problem on the downlink for wireless cellular networks. As a consequence of uncoordinated local scheduling decisions in classical networks, the base stations produce mutual uncontrolled interference on their co-channel users. This interference is of a variable nature, and is hardly predictable, which leads to suboptimal scheduling and power control decisions. While some works propose to introduce cooperation between base stations, in this work we propose instead to introduce a model of power variations, called trajectories in the powers space, to help each base station to predict the variations of other base stations powers. The trajectories are then updated using a Model Predictive Control (MPC) to adapt transmit powers according to a trade-off between inertia (to being predictable) and adaptation to fit with capacity needs. A Kalman filter is used for the interference prediction. In addition, the channel gains are also predicted, in order to anticipate channel fading states. This scheme can be seen as a dynamic distributed uncoordinated power control for multichannel transmission that fits the concept of self-optimised and self-organised wireless networks. By using the finite horizon MPC, the transmit powers are smoothly adapted to progressively leave the current trajectory toward the optimal trajectory. We formulate the optimisation problem as the minimisation of the utility function of the difference between the target powers and MPC predicted power values. The presented simulation results show that in dynamic channel conditions, the benefit of our approach is the reduction of the interference fluctuations, and as a consequence a more accurate interference prediction, which can further lead to a more efficient distributed scheduling, as well as the reduction of the overall power consumption.

6.2.3. BAN

In [26] we present a simple Body Area Network (BAN) platform that was built to monitor the performance of a marathon athlete all along the race, meeting real-time and QoS constraints, under good transmission conditions. Data collected during the event (packet loss, signal strength) allowed us to obtain a primary knowledge about the behavior of the radio transmissions between the different links in the network. The results of this experiment and ther important disparities observed between the links point out the need to improve the transmission strategy.

6.2.4. Network coding

One of the most powerful ways to achieve transmission reliability over wireless links is to employ efficient coding techniques. In [10] investigates the performance of a transmission over a relay channel where information is protected by two layers of coding. In the first layer, transmission reliability is ensured by fountain coding at the source. The second layer incorporates network coding at the relay node. Thus, fountain coded packets are re-encoded at the relay in order to increase packet diversity and reduce energy consumption. Performance of the transmission is measured by the total number of transmissions needed until the message is successfully decoded at the destination. We show through both analytical derivations and simulations that adding network coding capabilities at the relay optimizes system resource consumption. When the source uses a random linear fountain code, the proposed two layer encoding becomes more powerful as it reduces the transmission rate over the direct link between the source and the destination.

In [27] we study the deployment of fountain codes and network coding in a wireless sensor network (WSN). A WSN is composed of sensor nodes with restricted capacities: memory, energy and computational power. The nodes are usually randomly scattered across the monitored area and the environment may vary. In the presence of fading, outage and node failures, fountain codes are a promising solution to guaranty reliability and improve transmission robustness. The benefits of fountain codes are explored based on an event-driven WSNet simulator considering realistic implementation based on standard IEEE802.15.4. Fountain codes are rateless and capable of adapting their rate to the channel on the fly using a limited feedback channel. In this thesis, we highlight the benefits brought by fountain code in terms of energy consumption and transmission delay. In addition to the traditional transmission with fountain code, we propose in this thesis to study the network coding transmission scheme where nodes are allowed to process the information before forwarding it to their neighbors. By this means, we can say that packet diversity is exploited as each individual packet is unique and contains different representations of binary data. Redundancy is thus optimized since repetitions are avoided and replaced with diversified information. This can further lead to an overall improved performance in cooperative communication where nodes are allowed to assist in relaying packets from the source the destination. We highlight in this thesis the benefits of fountain code combined to network coding and show that it leads to a reduction in transmission delay and energy consumption. The latter is vital to the life duration of any wireless sensor network.

In [9] we tackles the problem of providing end to end reliable transmissions in a randomly deployed wireless sensor network. To this aim, we investigate the simultaneous use of gradient broadcast routing (for its inherent adaptability to any network topology and its changes), fountain codes (for their universal property) and intra-flow network coding (to introduce packet diversity in redundant copies). We present the impact of the proposed strategy on a realistic network. This work permits to highlight that, compared to basic gradient broadcast routing, the strategy not only improves the reliability and the delay in the network but also clearly increases its lifetime.

6.2.5. Vehicular networks

In [22] we study a hybrid propagation model For large-scale variations caused By vehicular traffic in small cells. we present a propagation model generating time series of large-scale power variations for small-cell radio links intersected by vehicular traffic. The model combines stochastic processing and geometric computation. For each road crossing a link, a two-state process parameterized by mobility statistics represents the obstruction status. When the status is set to obstructed, a fluctuation pattern is generated. Based on previously published measurements, both mobility statistics and time series results are validated through the comparison of respectively inter-obstruction duration distributions and outage probabilities. The proposed model avoids resource consuming iterative propagation prediction while providing realistic and frequency adaptive results.

In [21], we performed measurements of large-scale variations caused by vehicular traffic in small-cell. This paper presents and characterizes large-scale variations of received power generated by vehicular traffic crossing a radio link. Measurements in the 2 GHz band for several small-cell configurations involve various transmitter heights, link distances and urban densities. Observations showed that stronger losses up to 30 dB are due to medium to high vehicles. Lower vehicles have a smaller impact in links perpendicular to traffic, but amplitude variations and duration can reach larger values when the receiver is at cell radius limits.

6.2.6. security

In [18] we study Security Embedding on ultra wideband impulse radio(UWB-IR) Physical Layer. The main goal of this work is to incorporate security in an existing ultra wideband (UWB) network. We present an embedding method where a tag is added at the physical layer and superimposed to the UWB-impulse radio signal. The tag should be added in a transparent way so that guaranteeing compatibility with existing receivers ignoring the presence of the tag. We discuss technical details of the new embedding method. In addition, we discuss embedding strength and we analyze robustness performance. We demonstrate that the proposed embedding technique meets all the system design constraints.

In [11] we study Jamming in time-hopping ultrawide band (TH-UWB) Radio. With the great expansion of wireless communications, jamming becomes a real threat.We propose a new model to evaluate the robustness of a communication system to jamming. The model results in more scenarios to be considered ranging from the favorable case to the worst case. The model is applied to a TH-UWB radio. The performance of such a radio in presence of the different jamming scenarios is analyzed. We introduce a mitigation solution based on stream cipher that restricts the jamming problem of the TH-UWB communication to the more favorable case while preserving confidentiality.

6.2.7. Network Information Theory

Fundamental performance limits of multi-hop wireless transmissions are being investigated in [33] from a multiobjective perspective where transmission decisions (i.e. relay selection, scheduling or routing decision) modify the trade-off between capacity, reliability, end-to-end delay or network-wide energy consumption. In our previous work presented in the Inria research report RR-7799, Pareto-optimal performance bounds and network parameters have been derived for a 1-relay and 2-relay network within a MultiObjective(MO) performance evaluation framework. We show in this report that these bounds are tight since they can be reached by simple practical coding strategies performed by the source and the relays. Such strategies constitute achievable lower MO performance bounds on the real MO performance limits. More precisely, we adopt a coding strategy where the source transmits a random linear fountain code which is coupled to a network coding strategy performed by the relays. Two different network coding strategies are investigated. Practical performance bounds for both strategies are compared to the theoretical bound. We show that the theoretical bound is tight: generational distance between the practical and theoretical bound for the best strategy is only of 0.0042

In [19] we revisit the problem of non-cooperativ association of mobiles to access points using game theory. We consider in this paper games related to the association problem of mobiles to an access point. It consists of deciding to which access point to connect. We consider the choice between two access points or more, where the access decisions may depend on the number of mobiles connected to each one of the access points. We obtain new results using elementary tools in congestion and crowding games.

In [23] we study stochastic analysis of energy savings with sleep mode in Orthogonal Frequency-Division Multiple Access (OFDMA) wireless networks. The issue of energy efficiency in OFDMA wireless networks is discussed in this paper. Our interest is focused on the promising concept of base station sleep mode, introduced recently as a key feature in order to dramatically reduce network energy consumption. The proposed technical approach fully exploits the properties of stochastic geometry, where the number of active cells is reduced in a way that the outage probability, or equivalently the signal to interference plus noise distribution, remains the same. The optimal energy efficiency gains are then specified with the help of a simplified but yet realistic base station power consumption model. Furthermore, the authors extend their initial work by studying a non-singular path loss model in order to verify the validity of the analysis and finally, the impact on the achieved user capacity is investigated. In this context, the significant contribution of this paper is the evaluation of the theoretically optimal energy savings of sleep mode, with respect to the decisive role that the base station power profile plays.

6.3. Software Radio Programming Model

6.3.1. Virtual Radio Machine

In [14] we present a survey of existing prototypes dedicated to software defined radio. We propose a classification related to the architectural organization of the prototypes and provide some conclusions about the most promising architectures. This study should be useful for cognitive radio testbed designers who have to choose between many possible computing platforms. We also introduce a new cognitive radio testbed currently under construction and explain how this study have influenced the test-bed designers choices.

6.3.2. Embedded systems

In [13], we explore new area/throughput trade-offs for the Girault, Poupard and Stern authentication protocol (GPS). This authentication protocol was selected in the NESSIE competition and is even part of the standard ISO/IEC 9798. The originality of our work comes from the fact that we exploit a fixed key to increase the throughput. It leads us to implement GPS using the Chapman constant multiplier. This parallel implementation is 40 times faster but 10 times bigger than the reference serial one. We propose to serialize this multiplier to reduce its area at the cost of lower throughput. Our hybrid Chapman's multiplier is 8 times faster but only twice bigger than the reference. Results presented here allow designers to adapt the performance of GPS authentication to their hardware resources. The complete GPS prover side is also integrated in the network stack of the POW-WOW sensor which contains an Actel IGLOO AGL250 FPGA as a proof of concept.

The people involved in this axes also published in the computer science flied. For instance in [1] static vulnerability detection in java service-oriented components is studied. In [12] A lightweight Hash function family based on FCSRs is studied.

TREC Project-Team

6. New Results

6.1. Design and Performance Analysis of Wireless Networks

Participants: François Baccelli, Bartłomiej Błaszczyszyn, Chung Shue Chen, Miodrag Jovanović, Holger Paul Keeler, Mir Omid Haji Mirsadeghi, Frédéric Morlot, Tien Viet Nguyen.

CDMA/UMTS, Wireless LANs, ad hoc networks, IEEE 802.11, mesh networks, cognitive radio, Hiperlan, CSMA, TCP, MAC protocols, exponential back-off protocols, signal to interference ratio, coverage, capacity, transport capacity, admission and congestion control.

This axis bears on the analysis and the design of wireless access communication networks. Our contributions are organized in terms of network classes: cellular networks, wireless LANs and MANETs, VANETs. We also have a section on generic results that regard more general wireless networks. We are interested both in macroscopic models, which are particularly important for economic planning and in models allowing the definition and the optimization of protocols. Our approach combines several tools, queueing theory, point processes, stochastic geometry, random graphs, distributed control algorithms, self organization protocols.

6.1.1. Cellular Networks

The activity on cellular networks has several complementary facets ranging from performance evaluation to protocol design. The work is mainly based on strong collaborations with Alcatel-Lucent and Orange Labs.

6.1.1.1. Effect of Opportunistic Scheduling on the Quality of Service Perceived by the Users in OFDMA Cellular Networks

Our objective in [17] is to analyze the impact of fading and opportunistic scheduling on the quality of service perceived by the users in an Orthogonal Frequency Division Multiple Access (OFDMA) cellular network. To this end, assuming Markovian arrivals and departures of customers that transmit some given data volumes, as well as some temporal channel variability (fading), we study the mean throughput that the network offers to users in the long run of the system. Explicit formulas are obtained in the case of allocation policies, which may or may-not take advantage of the fading, called respectively opportunistic and non-opportunistic. The main practical results of the present work are the following. Firstly we evaluate for the non-opportunist allocation the degradation due to fading compared to Additive White Gaussian Noise (AWGN) (that is, a decrease of at least 13% of the throughput). Secondly, we evaluate the gain induced by the opportunistic allocation. In particular, when the traffic demand per cell exceeds some value (about 2 Mbits/s in our numerical example), the gain induced by opportunism compensates the degradation induced by fading compared to AWGN. Partial results were presented at ComNet in 2009 [61].

6.1.1.2. Impact of propagation-loss model on the geometry and performance of cellular networks

6.1.1.2.1. Impact of Shadowing on QoS

Shadowing is believed to degrade the quality of service in wireless cellular networks. In [18] we discovered a more subtle reality. Increasing variance of the lognormal shadowing tends to “separate” the strongest (serving BS) signal from all other signals — a phenomenon observed for heavy-tailed distributions and called “single big jump principle”. In consequence, in some cases, an increase of the variance of the shadowing can significantly reduce the mean interference factor and improve some QoS metrics in interference limited systems. We exemplify this phenomenon, similar to stochastic resonance and related to the “single big jump principle” of the heavy-tailed log-nornal distribution, studying the blocking probability in regular, hexagonal networks in a semi-analytic manner, using a spatial version of the Erlang’s loss formula combined with Kaufman-Roberts algorithm.

6.1.1.2.2. Using Poisson processes to model lattice cellular networks

In [51] we mathematically proved that a large spatially homogeneous (arbitrary, including hexagonal) network is perceived by a typical user as an equivalent (infinite) Poisson network, provided shadowing is strong enough. This justifies an almost ubiquitous Poisson assumption made in the stochastic-analytic approach to study of the quality of user-service in cellular networks.

6.1.1.2.3. Linear-Regression Estimation of the Propagation-Loss Parameters Using Mobiles' Measurements

In [35] we proposed a new linear-regression model for the estimation of the path-loss exponent and the parameters of the shadowing from the propagation-loss data collected by the mobiles with respect to their serving base stations. The model is based on the aforementioned Poisson convergence result.

6.1.1.3. Quality of Real-Time Streaming in Wireless Cellular Networks

In [50] we present a new stochastic service model with service capacity sharing and interruptions, meant to be useful for the performance evaluation and dimensioning of wireless cellular networks offering real-time streaming, like e.g. mobile TV. Our general model takes into account Markovian, multi-class process of call arrivals, arbitrary streaming time distribution, and allows for a general service (outage) policy saying which users are temporarily denied the service due to insufficient service capacity. Using Palm theory formalism, we develop expressions for several important characteristics of this model, including mean time spent in outage and mean number of outage incidents for a typical user of a given class. We also propose some natural class of least-effort-served-first service policies, for which the aforementioned expressions can be efficiently evaluated on the basis of the Fourier analysis of Poisson process. Last but not least, we show how our model can be used to analyse the quality of real-time streaming in 3GPP Long Term Evolution (LTE) cellular networks. We identify and evaluate an optimal and a fair service policy, the latter being suggested by LTE implementations, as well as propose some intermediate policies which allow to solve the optimality/fairness trade-off caused by unequal user radio-channel conditions.

6.1.1.4. Theoretically Feasible QoS in a MIMO Cellular Network Compared to the Practical LTE Performance

Our goal in [39] is to build a global analytical approach for the evaluation of the quality of service perceived by the users in wireless cellular networks which is calibrated in some reference cases. To do so, a model accounting for interference in a MIMO cellular system is firstly described. An explicit expression of users bit-rates theoretically feasible from the information theory point of view is then deduced. The comparison between these bit-rates and practical LTE performance permits to obtain the progress margins for potential evolution of the technology. Moreover, it leads to an analytical approximate expression of the system performance which is calibrated with the practical one. This expression is the keystone of a global analytical approach for the evaluation of the QoS perceived by the users in the long run of users arrivals and departures in the network. We illustrate our approach by calculating the users QoS as function of the cell radius in different mobility and interference cancellation scenarios.

6.1.1.5. Self-Optimization of Radio Resources in Cellular Networks

In [19], we surveyed the mathematical and algorithmic tools for the self-optimization of mobile cellular networks based on Gibbs' sampler. This technique allows for the joint optimization of radio resources in heterogeneous cellular networks made of a juxtaposition of macro and small cells. It can be implemented in a distributed way and nevertheless achieves minimal system-wide potential delay. Results show that it is effective in both throughput and energy efficiency.

Three patents were filed on this line of thought under the Inria/Alcatel–Lucent joint laboratory.

6.1.1.6. Coverage in Cellular Networks

Cellular networks are in a major transition from a carefully planned set of large tower-mounted base-stations (BSs) to an irregular deployment of heterogeneous infrastructure elements that often additionally includes micro, pico, and femtocells, as well as distributed antennas. In a collaboration with H. Dhillon, J. Andrews and R. Ganti [UT Austin, USA] [20], we developed a model for a downlink heterogeneous cellular network (HCN) consisting of K tiers of randomly located BSs, where each tier may differ in terms of average transmit power, supported data rate and BS density. Assuming a mobile user connects to the strongest candidate BS, the resulting Signal-to-Interference-plus-Noise-Ratio (SINR) is greater than 1 when in coverage, Rayleigh fading, we derived an expression for the probability of coverage (equivalently outage) over the entire network under both open and closed access. One interesting observation for interference-limited open access networks is that at a given SINR, adding more tiers and/or BSs neither increases nor decreases the probability of coverage or outage when all the tiers have the same SINR threshold.

6.1.2. Mobile Ad Hoc Networks

A MANET is made of mobile nodes which are at the same time terminals and routers, connected by wireless links, the union of which forms an arbitrary topology. The nodes are free to move randomly and organize themselves arbitrarily. Important issues in such a scenario are connectivity, medium access (MAC), routing and stability. This year, we worked on a game theoretic view of Spatial Aloha in collaboration with E. Altman and M.K. Hanawal [Inria MAESTRO] [22] This line of though is currently continued with Chandramani Singh. We also compared the performance of spatial Aloha to CSMA.

6.1.2.1. Improvement of CSMA/CA's Spatial Reuse

The most popular medium access mechanism for such ad hoc networks is CSMA/CA with RTS/CTS. In CSMA-like mechanisms, spatial reuse is achieved by implementing energy based guard zones. In a collaboration with Qualcomm [12], we considered the problem of simultaneously scheduling the maximum number of links that can achieve a given signal to interference ratio (SIR). Using tools from stochastic geometry, we studied and maximized the medium access probability of a typical link. Our contributions are two-fold: (i) We showed that a simple modification to the RTS/CTS mechanism, viz., changing the receiver yield decision from an energy-level guard zone to an SIR guard zone, leads to performance gains; and (ii) We showed that this combined with a simple modification to the transmit power level – setting it to be inversely proportional to the square root of the link gain – leads to significant improvements in network throughput. Further, this simple power-level choice is no worse than a factor of two away from optimal over the class of all "local" power level selection strategies for fading channels, and further is optimal in the non-fading case. The analysis relies on an extension of the Matérn hard core point process which allows us to quantify both these SIR guard zones and this power control mechanism.

6.1.2.2. Comparison of the maximal spatial throughput of Aloha and CSMA in Wireless multihop Ad-Hoc Networks

In [46] this paper we compare the spatial throughput of Aloha and Carrier Sense Multiple Access (CSMA) in Wireless multihop Ad-Hoc Networks. In other words we evaluate the gain offered by carrier sensing (CSMA) over the pure statiscal collision avoidance which is the basis of Aloha. We use a Signal-to-Interference-and-Noise Ratio (SINR) model where a transmission is assumed to be successful when the SINR is larger than a given threshold. Regarding channel conditions, we consider both standard Rayleigh and negligible fading. For slotted and non-slotted Aloha, we use analytical models as well as simulations to study the density of successful transmissions in the network. As it is very difficult to build precise models for CSMA, we use only simulations to compute the performances of this protocol. We compare the two Aloha versions and CSMA on a fair basis, i.e. when they are optimized to maximize the density of successful transmissions. For slotted Aloha, the key optimization parameter is the medium access probability, for non-slotted Aloha we tune the mean back-off time, whereas for CSMA it is the carrier sense threshold that is adjusted. Our study shows that CSMA always outperforms slotted Aloha, which in turn outperforms its non-slotted version.

6.1.2.3. Stochastic Analytic Evaluation of End-to-End Performance of Linear Nearest Neighbour Routing in MANETs with Aloha

Planar Poisson models with the Aloha medium access scheme have already proved to be very useful in studies of mobile ad-hoc networks (MANETs). However, it seems difficult to quantitatively study the performances of end-to-end routing in these models. In order to tackle this problem, in [52], we study a linear stationary route embedded in an independent planar field of interfering nodes. We consider this route as an idealization of a "typical" route in a MANET obtained by some routing mechanism. Such a decoupling allows us to obtain many numerically tractable expressions for local and mean end-to-end delays and the speed of packet progression, assuming slotted Aloha MAC and the Signal-to-Interference-and-Noise Ratio (SINR) capture condition, with the usual power-law path loss model and Rayleigh fading. These expressions show how the network performance depends on the tuning of Aloha and routing parameters and on the external noise level. In particular we show a need for a well-tuned lattice structure of fixed relaying nodes, which helps to relay packets on long random routes in the presence of a non-negligible noise. We also consider a Poisson-line MANET model, in which nodes are located on roads forming a Poisson-line process. In this case our linear route is rigorously (in the sense of Palm theory) the typical route in this Poisson-line MANET.

6.1.3. Vehicular Ad-Hoc Networks (VANETs)

Vehicular Ad Hoc NETworks (VANETs) are special cases of MANETs where the network is formed between vehicles. VANETs are today the most promising civilian application for MANETs and they are likely to revolutionize our traveling habits by increasing safety on the road while providing value added services.

6.1.3.1. Point-to-Point, Emergency and Broadcast Communications

Our aim in [36] is to analyze the Aloha medium access (MAC) scheme in one-dimensional, linear networks, which might be an appropriate assumption for VANETs. The locations of the vehicles are assumed to follow a homegeneous Poisson point process. Assuming powerlaw mean path-loss and independent point-to-point fading we study performance metrics based on the signal-over-interference and noise ratio (SINR). In contrast to previous studies where the receivers are at a fixed distance from the transmitter, we assume here that the receivers are the nearest neighbors of the transmitters in the Poisson process and in a given direction. We derive closed formulas for the capture probability and for the density of progress of a packet sent by a given node. We compute the mean delay to send a packet transmitted at each slot until successful reception. We also evaluate an upper bound to discover the neighborhood within a given space interval. We show that we can include noise in the previous models.

6.1.4. Cognitive Radio Networks

We wrote a survey [26] on the probabilistic framework which can be used to model and analyze cognitive radio networks using various classes of MAC protocols (including carrier sensing based multiple access schemes and Aloha schemes). For each model, analytical results were derived for important performance metrics. This leads to a quantification of the interplay between primary and secondary users in such networks.

6.2. Network Dynamics

Participants: Abir Benabid, Julieta Bollati, Anne Bouillard, Ana Bušić, Emilie Coupechoux, Nadir Farhi.

Queueing network, stability, inversion formula, probing, estimator, product-form, insensitivity, markov decision, max-plus algebra, network calculus.

6.2.1. Network Calculus

Network calculus is a theory that aims at computing deterministic performance guarantees in communication networks. This theory is based on the (min,plus) algebra. Flows are modeled by an *arrival curve* that upper-bounds the amount of data that can arrive during any interval, and network elements are modeled by a *service curve* that gives a lower bound on the amount of service offered to the flows crossing that element. Worst-case performances are then derived by combining these curves.

6.2.1.1. Performance bounds in FIFO tandem networks

In cooperation with Giovanni Stea [University of Pisa, Italy], we present in [31] algorithms to compute worst-case performance upper bounds when the service policy is FIFO, using linear programming. Linear programming leads to tight bounds; however, the computation corst is too high for reasonable-size networks. We then develop approximate solution schemes to find both upper and lower delay bounds on the worst-case delay. Both of them only require to solve just one LP problem, and they produce bounds which are generally more accurate than those found in the literature. Finally, we have a conjecture on what sould be the worst-case trajectory under usual assumptions.

6.2.1.2. *Feed-forward networks with wormhole routing discipline*

In collaboration with Bruno Gaujal [Inria Rhone Alpes] and Nadir Farhi [IFFSTAR] we are working on a model of performance bound calculus on feed-forward networks where data packets are routed under wormhole routing discipline. We are interested in determining maximum end-to-end delays and backlogs for packets going from a source node to a destination node, through a given virtual path in the network. Our objective is to give a “network calculus” approach to calculate the performance bounds. For this, we propose a new concept of curves that we call *packet curves*. The curves permit to model constraints on packet lengths for data flows, when the lengths are allowed to be different. We used this new concept to propose an approach for calculating residual services for data flows served under non preemptive service disciplines. This notion also enabled us to differentiate different classes of service policies: those that are based on a packet count (like round-robin and its generalized version), where the packet curve will be useful to tighten the bounds computed, and those that are based on the amount of data served (FIFO, priorities), where it won’t be useful. These results have been presented at Valuetools (invited paper, [29]).

6.2.1.3. *Using arrival curves for detecting anomalies in a network*

In cooperation with Aurore Junier [Inria/IRISA] and Benoît Ronot [Alcatel-Lucent], we present an on-line algorithm that performs a flow of messages analysis. More precisely, it is able to highlight hidden abnormal behaviors that existing network management methods would not detect. Our algorithm uses the notion of constraint curves, introduced in the Network Calculus theory, defining successive time windows that bound the flow. The advantage of this algorithm is that it can be performed online, and in a second version has different levels of precision. This work has been presented in [30] and a patent [57] has been submitted.

6.2.1.4. *Min,plus algorithms for fast weak-KAM integrators*

In cooperation with Erwan Faou [IPSO-Inria Rennes, DMA-ENS] and Maxime Zavidovique [Paris 6]. We consider a numerical scheme for Hamilton-Jacobi equations based on a direct discretization of the Lax-Oleinik semi-group. We prove that this method is convergent with respect to the time and space stepsizes provided the solution is Lipschitz, and give an error estimate. Moreover, we prove that the numerical scheme is a geometric integrator satisfying a discrete weak-KAM theorem which allows to control its long time behavior. Taking advantage of a fast algorithm for computing min-plus convolutions based on the decomposition of the function into concave and convex parts, we show that the numerical scheme can be implemented in a very efficient way. The results can be found in [49].

6.2.2. *Perfect Sampling of Queueing Systems*

Propp and Wilson introduced in 1996 a perfect sampling algorithm that uses coupling arguments to give an unbiased sample from the stationary distribution of a Markov chain on a finite state space $\mathcal{X}$. In the general case, the algorithm starts trajectories from all $x \in \mathcal{X}$ at some time in the past until time $t = 0$. If the final state is the same for all trajectories, then the chain has coupled and the final state has the stationary distribution of the Markov chain. Otherwise, the simulations are started further in the past. This technique is very efficient if all the events in the system have appropriate monotonicity properties. However, in the general (non-monotone) case, this technique requires that one consider the whole state space, which limits its application only to chains with a state space of small cardinality.

6.2.2.1. *Piecewise Homogeneous Events*

In collaboration with Bruno Gaujal [Inria Grenoble - Rhone-Alpes], we proposed in [15] a new approach for the general case that only needs to consider two trajectories. Instead of the original chain, we used two bounding processes (envelopes) and we showed that, whenever they couple, one obtains a sample under the stationary distribution of the original chain. We showed that this new approach is particularly effective when the state space can be partitioned into pieces where envelopes can be easily computed. We further showed that most Markovian queueing networks have this property and we propose efficient algorithms for some of them.

The envelope technique has been implemented in a software tool PSI2 (see Section 5.2).

6.2.2.2. Perfect Sampling of Networks with Finite and Infinite Capacity Queues

In [33], we consider open Jackson queueing networks with mixed finite and infinite buffers and analyze the efficiency of sampling from their exact stationary distribution. We show that perfect sampling is possible, although the underlying Markov chain has a large or even infinite state space. The main idea is to use a Jackson network with infinite buffers (that has a product form stationary distribution) to bound the number of initial conditions to be considered in the coupling from the past scheme. We also provide bounds on the sampling time of this new perfect sampling algorithm under hyper-stability conditions (to be defined in the paper) for each queue. These bounds show that the new algorithm is considerably more efficient than existing perfect samplers even in the case where all queues are finite. We illustrate this efficiency through numerical experiments.

6.2.3. Markov Chains and Markov Decision Processes

Solving Markov chains is in general difficult if the state space of the chain is very large (or infinite) and lacking a simple repeating structure. One alternative to solving such chains is to construct models that are simple to analyze and provide bounds for a reward function of interest. The bounds can be established by using different qualitative properties, such as stochastic monotonicity, convexity, submodularity, etc. In the case of Markov decision processes, similar properties can be used to show that the optimal policy has some desired structure (e.g. the critical level policies).

6.2.3.1. Stochastic Monotonicity

In collaboration with Jean-Michel Fourneau [PRiSM, Université de Versailles Saint-Quentin] we consider two different applications of stochastic monotonicity in performance evaluation of networks [14]. In the first one, we assume that a Markov chain of the model depends on a parameter that can be estimated only up to a certain level and we have only an interval that contains the exact value of the parameter. Instead of taking an approximated value for the unknown parameter, we show how we can use the monotonicity properties of the Markov chain to take into account the error bound from the measurements. In the second application, we consider a well known approximation method: the decomposition into submodels. In such an approach, models of complex networks are decomposed into submodels whose results are then used as parameters for the next submodel in an iterative computation. One obtains a fixed point system which is solved numerically. In general, we have neither an existence proof of the solution of the fixed point system nor a convergence proof of the iterative algorithm. Here we show how stochastic monotonicity can be used to answer these questions. Furthermore, monotonicity properties can also help to derive more efficient algorithms to solve fixed point systems.

6.2.3.2. Markov Reward Processes and Aggregation

In a joint work with I.M. H. Vliegen [University of Twente, The Netherlands] and A. Scheller-Wolf [Carnegie Mellon University, USA] [16], we presented a new bounding method for Markov chains inspired by Markov reward theory: Our method constructs bounds by redirecting selected sets of transitions, facilitating an intuitive interpretation of the modifications of the original system. We show that our method is compatible with strong aggregation of Markov chains; thus we can obtain bounds for an initial chain by analyzing a much smaller chain. We illustrated our method by using it to prove monotonicity results and bounds for assemble-to-order systems.

6.2.3.3. Bounded State Space Truncation

Markov chain modeling often suffers from the curse of dimensionality problems and many approximation schemes have been proposed in the literature that include state-space truncation. Estimating the accuracy of such methods is difficult and the resulting approximations can be far from the exact solution. Censored Markov chains (CMC) allow to represent the conditional behavior of a system within a subset of observed states and provide a theoretical framework to study state-space truncation. However, the transition matrix of a CMC is in general hard to compute. Dayar et al. (2006) proposed DPY algorithm, that computes a stochastic bound for a CMC, using only partial knowledge of the original chain. In [32], we prove that DPY is optimal for the information they take into account. We also show how some additional knowledge on the chain can improve stochastic bounds for CMC.

6.2.4. Dynamic Systems with Local Interactions

Dynamic systems with local interactions can be used to model problems in distributed computing: gathering a global information by exchanging only local information. The challenge is two-fold: first, it is impossible to centralize the information (cells are indistinguishable); second, the cells contain only a limited information (represented by a finite alphabet $\mathcal{A}$; $\mathcal{A} = \{0, 1\}$ in our case). Two natural instantiations of dynamical systems are considered, one with synchronous updates of the cells, and one with asynchronous updates. In the first case, time is discrete, all cells are updated at each time step, and the model is known as a *Probabilistic Cellular Automaton (PCA)* (e.g. Dobrushin, R., Kryukov, V., Toom, A.: *Stochastic cellular systems: ergodicity, memory, morphogenesis*, 1990). In the second case, time is continuous, cells are updated at random instants, at most one cell is updated at any given time, and the model is known as a (finite range) *Interacting Particle System (IPS)* (e.g. Liggett, T.M.: *Interacting particle systems*, 2005).

6.2.4.1. Density Classification on Infinite Lattices and Trees

In a joint work with N. Fatès [Inria Nancy – Grand-Est], J. Mairesse and I. Marcovici [LIAFA, CNRS and Université Paris 7] [43] we consider an infinite graph with nodes initially labeled by independent Bernoulli random variables of parameter p. We address the density classification problem, that is, we want to design a (probabilistic or deterministic) cellular automaton or a finite-range interacting particle system that evolves on this graph and decides whether p is smaller or larger than $1/2$. Precisely, the trajectories should converge (weakly) to the uniform configuration with only $0's$ if $p < 1/2$, and only $1's$ if $p > 1/2$. We present solutions to that problem on $\mathbb{Z}^d$, for any $d \geq 2$, and on the regular infinite trees. For $\mathbb{Z}$, we propose some candidates that we back up with numerical simulations.

6.3. Economics of Networks

Participants: François Baccelli, Emilie Coupechoux, Marc Lelarge.

6.3.1. Diffusion and Cascading Behavior in Random Networks

The spread of new ideas, behaviors or technologies has been extensively studied using epidemic models. In [25], we consider a model of diffusion where the individuals' behavior is the result of a strategic choice. We study a simple coordination game with binary choice and give a condition for a new action to become widespread in a random network. We also analyze the possible equilibria of this game and identify conditions for the coexistence of both strategies in large connected sets. Finally we look at how can firms use social networks to promote their goals with limited information.

Our results differ strongly from the one derived with epidemic models. In particular, we show that connectivity plays an ambiguous role: while it allows the diffusion to spread, when the network is highly connected, the diffusion is also limited by high-degree nodes which are very stable. In the case of a sparse random network of interacting agents, we compute the contagion threshold for a general diffusion model and show the existence of (continuous and discontinuous) phase transitions. We also compute the minimal size of a seed of new adopters in order to trigger a global cascade if these new adopters can only be sampled without any information on the graph. We show that this minimal size has a non-trivial behavior as a function of the connectivity. Our analysis extends methods developed in the random graphs literature based on the properties of empirical distributions of independent random variables, and leads to simple proofs.

6.3.2. Coordination in Network Security Games: a Monotone Comparative Statics Approach

Malicious softwares or malwares for short have become a major security threat. While originating in criminal behavior, their impact are also influenced by the decisions of legitimate end users. Getting agents in the Internet, and in networks in general, to invest in and deploy security features and protocols is a challenge, in particular because of economic reasons arising from the presence of network externalities. In [24], [42], we focus on the question of incentive alignment for agents of a large network towards a better security. We start with an economic model for a single agent, that determines the optimal amount to invest in protection. The model takes into account the vulnerability of the agent to a security breach and the potential loss if a security breach occurs. We derive conditions on the quality of the protection to ensure that the optimal amount spent on security is an increasing function of the agent's vulnerability and potential loss. We also show that for a large class of risks, only a small fraction of the expected loss should be invested. Building on these results, we study a network of interconnected agents subject to epidemic risks. We derive conditions to ensure that the incentives of all agents are aligned towards a better security. When agents are strategic, we show that security investments are always socially inefficient due to the network externalities. Moreover alignment of incentives typically implies a coordination problem, leading to an equilibrium with a very high price of anarchy.

6.4. Point Processes, Stochastic Geometry and Random Geometric Graphs

Participants: François Baccelli, Bartłomiej Błaszczyszyn, Pierre Brémaud, Kumar Gaurav, Mir Omid Haji Mirsadeghi.

stochastic geometry, point process, shot-noise, Boolean model, random tessellation, percolation, stochastic comparison

6.4.1. Modeling, comparison and impact of spatial irregularity of point processes on coverage, percolation, and other characteristics of random geometric models

We develop a general approach for comparison of clustering properties of point processes. It is funded on some basic observations allowing to consider void probabilities and moment measures as two complementary tools for capturing clustering phenomena in point processes. As expected, smaller values of these characteristics indicate less clustering. Also, various global and local functionals of random geometric models driven by point processes admit more or less explicit bounds involving the void probabilities and moment measures, thus allowing to study the impact of clustering of the underlying point process. When stronger tools are needed, dcx ordering of point processes happens to be an appropriate choice, as well as the notion of (positive or negative) association, when comparison to the Poisson point process is concerned. The whole approach has been worked out in a series of papers [62], [63], [64], [65]. This year we have prepared revisions of the two latter ones, from which [65] is now accepted for the publication in Adv. Appl. Probab. We have also prepared a review article [53] for *Lecture Notes in Mathematics*, Springer.

6.4.1.1. AB random geometric graphs

We investigated percolation in the AB Poisson-Boolean model in d-dimensional Euclidean space, and asymptotic properties of AB random geometric graphs on Poisson points in $[0,1]^d$. The AB random geometric graph we studied is a generalization to the continuum of a bi-partite graph called the AB percolation model on discrete lattices. Such an extension is motivated by applications to secure communication networks and frequency division duplex networks. The AB Poisson Boolean model is defined as a bi-partite graph on two independent Poisson point processes of intensities λ and μ in the d-dimensional Euclidean space in the same manner as the usual Boolean model with a radius r. We showed existence of AB percolation for all $d \geq 2$, and derived bounds for a critical intensity. Further, in $d = 2$, we characterize a critical intensity. The set-up for AB random geometric graphs is to construct a bi-partite graph on two independent Poisson point process of intensities n and cnin the unit cube. We provided almost sure asymptotic bounds for the connectivity threshold for all $c > 0$ and a suitable choice of radius cut-off functions $r_n(c)$. Further for $c < c_0$, we derived a weak law result for the largest nearest neighbor radius. This work appeared in [27].

6.4.2. Random Packing Models

Random packing models (RPM) are point processes (p.p.s) where points which "contend" with each other cannot be simultaneously present. These p.p.s play an important role in many studies in physics, chemistry, material science, forestry and geology. For example, in microscopic physics, chemistry and material science, RPMs can be used to describe systems with hard-core interactions. Applications of this type range from reactions on polymer chains, chemisorption on a single-crystal surface, to absorption in colloidial systems. In these models, each point (molecule, particle,$\cdots$) in the system occupies some space, and two points with overlapping occupied space contend with each other. Another example is the study of seismic and forestry data patterns, where RPMs are used as a reference model for the data set under consideration. In wireless communications, RPMs can be used to model the users simultaneously accessing the medium in a wireless network using Carrier Sensing Medium Access (CSMA). In this context, each point (node, user, transmitter,$\cdots$) does not occupy space but instead generates interference to other points in the network. Two points contend with each other if either of them generates too much interference to the other. Motivated by this kind of application, we studied in [66] the generating functionals of several models of random packing processes: the classical Matérn hard-core model; its extensions, the k-Matérn models and the ∞-Matérn model, which is an example of random sequential packing process. The main new results are: 1) A sufficient condition for the ∞-Matérn model to be well-defined (unlike the other two, the ∞-Matérn model may not be well-defined on unbounded space); 2) the generating functional of the resulting point process which is given for each of the three models as the solution of a differential equation; 3) series representation and bounds on the generating functional of the packing models; 4) moment measures and other useful properties of the considered packing models which are derived from their generating functionals.

6.4.3. Extremal and Additive Matérn Point Processes

In the simplest Matérn point processes, one retains certain points of a Poisson point process in such a way that no pairs of points are at distance less than a threshold. This condition can be reinterpreted as a threshold condition on an extremal shot–noise field associated with the Poisson point process. In a joint work with P. Bermolen [Universidad de la República, Montevideo, Uruguay] [60], we studied extensions of Matérn point processes where one retains points that satisfy a threshold condition based on an *additive* shot–noise field of the Poisson point process. We provide an analytical characterization of the intensity of this class of point processes and we compare the packing obtained by the extremal and additive schemes and certain combinations thereof.

6.4.4. Spatial Birth and Death Point Processes

In collaboration with F. Mathieu [Inria GANG] and Ilkka Norros [VTT, Finland], we continued studying a new spatial birth and death point process model where the death rate is a shot noise of the point configuration. We showed that the spatial point process describing the steady state exhibits repulsion. We studied two asymptotic regimes: the fluid regime and the hard–core regime. We derived closed form expressions for the mean (and in some cases the law) of the latency of points as well as for the spatial density of points in the steady state of each regime. A paper on the matter will be presented at Infocom 13.

6.4.5. A population model based on a Poisson line tessellation

In [44], we introduce a new population model. Taking the geometry of cities into account by adding roads, we build a Cox process driven by a Poisson line tessellation. We perform several shot-noise computations according to various generalizations of our original process. This allows us to derive analytical formulas for the uplink coverage probability in each case.

6.4.6. Information Theory and Stochastic Geometry

In a joint work with V. Anantharam [UC Berkeley], we study the Shannon regime for the random displacement of stationary point processes. We currently investigate Multiple Access Channels.

6.4.7. Navigation on Point Processes and Graphs

The thesis of Mir Omid Mirsadeghi [6] studied optimal navigations in wireless networks in terms of first passage percolation on some space-time SINR graph. It established both “positive” and “negative” results on the associated percolation delay rate (delay per unit of Euclidean distance, also called time constant in the classical terminology of percolation). The latter determines the asymptotics of the minimum delay required by a packet to progress from a source node to a destination node when the Euclidean distance between the two tends to infinity. The main negative result states that the percolation delay rate is infinite on the random graph associated with a Poisson point process under natural assumptions on the wireless channels. The main positive result states that when adding a periodic node infrastructure of arbitrarily small intensity to the Poisson point process, the percolation delay rate is positive and finite.

A new direction of research was initiated aiming at defining a new class of measures on a point process which are invariant under the action of a navigation on this point process. This class of measures has properties similar to Palm measures of stationary point processes; but they cannot be defined in the classical framework of Palm measures.

6.5. Random Graphs and Combinatorial Optimization

Participants: Emilie Coupechoux, Kumar Gaurav, Mathieu Leconte, Marc Lelarge.

random graphs, combinatorial optimization, local weak convergence, diffusion, network games.

6.5.1. Matchings in infinite graphs

In [13] with Charles Bordenave [CNRS-Université de Toulouse] and Justin Salez [Université Paris 7], we proved that for any sequence of (deterministic or random) graphs converging locally, the corresponding sequence of normalized matching numbers converges, and this limit depends only on the limit of the graph sequence. In the particular case where this limit is a unimodular Galton Watson tree, we were able to compute explicitly the value for the limit of the sequence of (normalized) matching numbers. This leads to an explicit formula that considerably extends the well-known one by Karp and Sipser for Erdős-Rényi random graphs.

We considered a natural family of Gibbs distributions over matchings on a finite graph, parameterized by a single positive number called the temperature. The correlation decay technique can be applied for the analysis of matchings at positive temperature and allowed us to establish the weak convergence of the Gibbs marginal as the underlying graph converges locally. However for the zero temperature problem (i.e. maximum matchings), we showed that there is no correlation decay even in very simple cases. By using a complex temperature and a half-plane property due to Heilmann and Lieb, we were able to let the temperature tend to zero and obtained a limit theorem for the asymptotic size of a maximum matching in the graph sequence.

6.5.2. Convergence of Multivariate Belief Propagation, with Applications to Cuckoo Hashing and Load Balancing

In [58], with Laurent Massoulié [Inria-MSR], we extend the results obtained previously on the asymptotic size of maximum matchings in random graphs converging locally to Galton-Watson trees to so-called capacitated b-matchings (with non-unitary capacity at vertices as well as constraints on individual edges). Compared to the matching case, this involves studying the convergence of a message passing algorithms which transmits vectors instead of single real numbers. We also look further into an application of these results to large multiple-choice hashtables. In particular, cuckoo hashing is a popular and simple way to build a hashtable where each item is only allowed to be assigned keys within a predetermined, random subset of all keys. In this context, it is important to determine the load threshold under which cuckoo hashing will succeed with high probability in building such a hashtable. The results on the density of maximum capacitated b-matchings allow to determine this threshold.

6.5.3. A new approach to the orientation of random hypergraphs

A h-uniform hypergraph $H=(V,E)$ is called (l,k)-orientable if there exists an assignment of each hyperedge e to exactly l of its vertices such that no vertex is assigned more than k hyperedges. Let $H_{n,m,h}$ be a hypergraph, drawn uniformly at random from the set of all h-uniform hypergraphs with n vertices and m edges. In [41], we determine the threshold of the existence of a (l,k)-orientation of $H_{n,m,h}$ for $k \geq 1$ and $h > l \geq 1$, extending recent results motivated by applications such as cuckoo hashing or load balancing with guaranteed maximum load. Our proof combines the local weak convergence of sparse graphs and a careful analysis of a Gibbs measure on spanning subgraphs with degree constraints. It allows us to deal with a much broader class than the uniform hypergraphs.

6.5.4. Bipartite graph structures for efficient balancing of heterogeneous loads

In [40], with Laurent Massoulié [Inria-MSR], we look into another application of the results on the asymptotic maximum size of b-matchings to large scale distributed content service platforms, such as peer-to-peer video-on-demand systems. In this context, the density of maximum b-matchings corresponds to the maximum fraction of simultaneously satisfiable requests, when the service resources are limited and each server can only handle requests for a predetermined subset of the contents which it has stored in memory. An important design aspect of such systems is the content placement strategy onto the servers depending on the estimated content popularities; the results obtained allow to characterize the efficiency of such placement strategies and the optimal strategies in the limit of large storage capacity at servers are determined.

6.5.5. Flooding in Weighted Random Graphs

In a joint work [8] with Hamed Amini [EPFL] and Moez Draief [Imperial College London], we studied the impact of the edge weights on distances in diluted random graphs. We interpret these weights as delays, and take them as i.i.d exponential random variables. We analyzed the edge flooding time defined as the minimum time needed to reach all nodes from one uniformly chosen node, and the edge diameter corresponding to the worst case edge flooding time. Under some regularity conditions on the degree sequence of the random graph, we showed that these quantities grow as the logarithm of n, when the size of the graph n tends to infinity. We also derived the exact value for the prefactors.

These allowed us to analyze an asynchronous randomized broadcast algorithm for random regular graphs. Our results show that the asynchronous version of the algorithm performs better than its synchronized version: in the large size limit of the graph, it will reach the whole network faster even if the local dynamics are similar on average.

6.5.6. Upper deviations for split times of branching processes

In [9], upper deviation results are obtained for the split time of a supercritical continuous-time Markov branching process. More precisely, with Hamed Amini [EPFL], we establish the existence of logarithmic limits for the likelihood that the split times of the process are greater than an identified value and determine an expression for the limiting quantity. We also give an estimation for the lower deviation probability of the split times which shows that the scaling is completely different from the upper deviations.

6.5.7. Epidemics in random clustered networks

In [54], we study a model of random networks that has both a given degree distribution and a tunable clustering coefficient. We consider two types of growth processes on these graphs: diffusion and symmetric threshold model. The diffusion process is inspired from epidemic models. It is characterized by an infection probability, each neighbor transmitting the epidemic independently. In the symmetric threshold process, the interactions are still local but the propagation rule is governed by a threshold (that might vary among the different nodes). An interesting example of symmetric threshold process is the contagion process, which is inspired by a simple coordination game played on the network. Both types of processes have been used to model spread of new ideas, technologies, viruses or worms and results have been obtained for random graphs with no clustering. In this paper, we are able to analyze the impact of clustering on the growth processes. While clustering inhibits the diffusion process, its impact for the contagion process is more subtle and depends on the connectivity of the graph: in a low connectivity regime, clustering also inhibits the contagion, while in a high connectivity regime, clustering favors the appearance of global cascades but reduces their size. For both diffusion and symmetric threshold models, we characterize conditions under which global cascades are possible and compute their size explicitly, as a function of the degree distribution and the clustering coefficient. Our results are applied to regular or power-law graphs with exponential cutoff and shed new light on the impact of clustering.

6.5.8. Leveraging Side Observations in Stochastic Bandits

The paper [37] considers stochastic bandits with side observations, a model that accounts for both the exploration/exploitation dilemma and relationships between arms. In this setting, after pulling an arm i, the decision maker also observes the rewards for some other actions related to i. We will see that this model is suited to content recommendation in social networks, where users' reactions may be endorsed or not by their friends. We provide efficient algorithms based on upper confidence bounds (UCBs) to leverage this additional information and derive new bounds improving on standard regret guarantees. We also evaluate these policies in the context of movie recommendation in social networks: experiments on real datasets show substantial learning rate speedups ranging from 2.2x to 14x on dense networks.

6.5.9. Universality in Polytope Phase Transitions and Message Passing Algorithms

In [28], with Mohsen Bayati and Andrea Montanari [Stanford], we consider a class of nonlinear mappings F in R^N indexed by symmetric random matrices A in $R^{N\times N}$ with independent entries. Within spin glass theory, special cases of these mappings correspond to iterating the TAP equations and were studied by Erwin Bolthausen. Within information theory, they are known as 'approximate message passing' algorithms. We study the high-dimensional (large N) behavior of the iterates of F for polynomial functions F, and prove that it is universal, i.e. it depends only on the first two moments of the entries of A, under a subgaussian tail condition. As an application, we prove the universality of a certain phase transition arising in polytope geometry and compressed sensing. This solves -for a broad class of random projections- a conjecture by David Donoho and Jared Tanner.

6.5.10. Far-out Vertices In Weighted Repeated Configuration Model

In [34] we consider an edge-weighted uniform random graph with a given degree sequence (Repeated Configuration Model) which is a useful approximation for many real-world networks. It has been observed that the vertices which are separated from the rest of the graph by a distance exceeding certain threshold play an important role in determining some global properties of the graph like diameter, flooding time etc., in spite of being statistically rare. We give a convergence result for the distribution of the number of such far-out vertices. We also make a conjecture about how this relates to the longest edge of the minimal spanning tree on the graph under consideration.

TRISKELL Project-Team

6. New Results

6.1. Distributed models at runtime

In the last two years we have developed a new models@runtime approach, named Kevoree. It supports extensive architecture evolution at runtime and enables the design of eternal systems with a continuous design process. The Kevoree type model supports dynamic types redefinition, allowing for complete redesign of specifications and implementations while the system is running. Communication channels between components are themselves first class dynamic entities. By combining our component metamodel and a *models@runtime* approach we have developed implementations of Kevoree for a wide range of computation nodes, ranging from inexpensive embedded microcontrollers to large commercial cloud implementations. We have shown that **applications based on the Kevoree component model are able to reconfigure their architecture completely on the fly** several times per second [40] on computation nodes with very limited resources.

Using the Kevoree platform, we demonstrated the use of *models@runtime* for large-scale distributed systems. We have shown that the *models@runtime* approach is applicable to pervasive distributed systems, even with volatile networks and continuously changing topologies [41]. Using *ad hoc* distributed algorithms, architectural models are propagated reliably in spite of frequent loss of connectivity, and **reconfigurations of a distributed application are managed in a continuous consistent manner**. Using colored Petri nets to describe quantititative properties we are building a toolchain to estimate the time related properties of assemblies at runtime [51].

6.2. Real scale platform for dynamic tactical decision system

Since mid 2011 the Triskell team is designing and implementing the DAUM platform that integrates a large range of technologies, ranging from wireless low cost sensors to clouds made of rugged field miniservers. Our application use case is a tactical decision system designed in cooperation with a large firefighter department of 3,500 firefighters. This platform is being used as a real life testbed for our results on dynamic, continuous design of distributed pervasive systems. It is also used as a concrete cooperation support within the Marie Curie Initial Training Network *Relate*.

By combining *models@runtime* techniques and component-based techniques, we have shown how we can apply model driven engineering to design large-scale, distributed, heterogeneous and adaptive systems [40].

6.3. Software Language Engineering

With the growing interest in MDE, more and more models are used during a software development to capture various aspects (both functional and extra-functional). Therefore, explicitly identifying and analyzing these relationships becomes a real challenge during a model-based software development. To address this challenge, we proposed a **formal language that captures relations between modeled things in order to reason and communicate about modeling activities** [19].

More recently, we started to explore the necessary breakthrough in software languages to support a global software engineering. Consequently, we investigate MDE-based tools and methods in software language engineering (SLE) for the design and implementation of collaborative, interoperable and composable modeling languages [32], [31], [30].

6.4. Model Typing

In recent years, the Triskell team established a formal theory of model typing, considering models as first class entities when modeling in the large [8]. Model typing was initially developed to support the reuse of both metamodels and model transformations [21]. It is now becoming the cornerstone of the various established metamodeling operators to ensure structural and behavioral properties [85][43].

[8] Model typing goes beyond the typing of individual model elements to actually deal with the type of graphs of model elements

The series of work on model typing was initially developed in the context of Jim Steel's PhD, defended in 2008. Then, it has continuously evolved in the scheme of the Naouel Moha's post doctoral position and the Clément Guy's PhD thesis [43]. Recently, work on model typing had a very strong application to the field of optimizing compilers [18]. This is the result of a close collaboration between Inria and Colorado State University (CSU), involving two teams in MDE (the Triskell team at Inria and the SE group at CSU), and two teams in optimizing compilers (the CAIRN team at Inria and the Mélange group at CSU). This collaboration was partially funded by the Inria associated teams MoCAA and LRS.

6.5. Model Footprint / Pruning / Slicing

During the previous evaluation period, we have established various facilities to ease the metamodeling activity.

Model operations such as transformation and composition declare source metamodels that are usually larger than the set of concepts and relations actually used by the operation. We have proposed and validated a static operation analyzer to retrieve the metamodel footprint of the operation [46]. Then, we propose a conjunct use of model typing and metamodel pruning to ease the reuse of model transformations on instances of different metamodels [21].

In general, many operators consist into extracting a subset of a model according to a language-based specification. Model slicing is a model operation that consists in extracting a subset of a model. Because the creation of a new DSL implies the creation from scratch of a new model slicer, we proposed the Kompren language that models and generates model slicers for any DSL [70][66]. An extended version was recently published in SoSyM [14].

6.6. Model Composition

Triskell hence contributed to the software engineering community's effort to propose new ways of composing software from modeling elements, including for cross cutting concerns, that would unify the composition ideas behind Model Driven Engineering, Aspect Oriented Modeling, Software Product Lines etc [77]. Several research prototypes [9] have been built to provide new composition operators. In the Mickael Clavreul PhD [72], we define a framework to unify and classify existing model composition operator and ease the definition of new model composition operators. Theoretical basis to such a framework have been recently based on category theory in [48].

6.7. Model Variability

In the context of Aspects Oriented Modeling (AOM), one of the key challenge is the variability management leading to software product lines. Our work in this area has led to the involvement of the Triskell group in the ANR project MOVIDA, as well as in the OMG standardization process of the *Common Variability Language* where we developed a Kermeta-based implementation conforming to this future standard (called *kCVL*).

6.8. Testing software product lines

Nowadays, many applications are expected to run on a tremendous variety of execution environments. For example, network connection software must deliver the same functionalities on distinct physical platforms, which themselves run several distinct operating systems, with various applications and physical devices. Testing those applications is challenging as it is simply impossible to consider every possible environment configuration. We tackle this issue through the systematic selection of a subset of configurations for testing [45] and through model-based verification [37].

[9] http://www.kermeta.org/kompose/,http://www.kermeta.org/mdk/ModMap/

6.9. Testing service-oriented applications

The changes resulting from the evolution of Service Based Systems (SBSs) may degrade their design and quality of service (QoS) and may often cause the appearance of common poor solutions, called antipatterns. The automatic detection of antipatterns is thus important to assess the design and QoS of SBSs and ease their maintenance and evolution. Using our approach, we specify 10 well-known and common antipatterns, including Multi Service and Tiny Service, and we automatically generate their detection algorithms [50]. This work has received the best paper award at ICSOC 2012.

6.10. Testing aspect oriented programs

Aspect-oriented programming (AOP) promises better software quality through enhanced modularity.

Crosscutting concerns are encapsulated in separate units called aspects and are introduced at specific points in the base program at compile-time or runtime. However, aspect-oriented mechanisms also introduce new risks for reliability that must be tackled by specific testing techniques in order to fully benefit from the use of AOP. During the evaluation period, we proposed a series of work to analyze these new risks, let designers understand the interactions between the base and the aspects and test aspects. The major achievement is a **novel oracle to test the injection of aspects in a base program**. The oracle allows to capture new classes of errors that occur only in aspect-oriented programs. Its ability to capture these errors in a more efficient way than an object-oriented oracle (shorter test cases and written in less time), has been empirically demonstrated and was published in the Journal for Software Testing, Verification and Reliability [74].

6.11. Testing peer-to-peer systems

Peer-to-peer (P2P) is one of the major distributed platforms for many applications such as large data sharing and collaboration in social networks. However, building trustworthy P2P applications is difficult because they must be deployed on a large number of autonomous, volatile nodes, which may refuse to answer to some requests and even leave the system unexpectedly. This volatility of nodes is a common behavior in P2P systems and may be interpreted as a fault during tests (*i.e.*, failed node). In this context, we have developed a **novel framework and a methodology for testing P2P applications**. The framework is based on the individual control of nodes, allowing test cases to precisely control the volatility of nodes during their execution. We validated this framework through an experimentation on the FreePastry distributed hashtable. The experimentation tests the behavior of the system in different conditions of volatility and shows how the tests were able to detect complex implementation errors. This work, published in the Empirical Software Engineering journal [73], in collaboration with the ATLAS Inria team, is directly related to Triskell's goal to apply software engineering to distributed systems.

6.12. Testing the boundaries of a specific domain

The increasing use of domain-specific modeling to increase efficiency in modeling multiple concerns, increases the need to correctly formalize domain models. Domains are modeled as metamodels, which capture the domain's modeling spaces, *i.e.* the set of all models which structure conforms to the description specified in the metamodel. However, there is currently no systematic method to test that a metamodel captures all the correct models of the domain and no more. Our most recent contribution to testing focuses on the **automatic selection of models in the modeling space captured by a metamodel**. We adapt metaheuristic search to generate a set that covers as many representative situations as possible, while staying as small as possible. This work was published in the International Conference on Software Testing, verification and validation [27].

6.13. Testing interactive systems

While model-based design of interactive systems is moving from pure event-based models of WIMP interactions to stateful models of post-WIMP interactions, model-based test generation techniques for HCI currently consider only WIMP interaction testing. We proposed an original model-based test generation technique, which aims at providing test cases to test post-WIMP behavior (*e.g.* multi-touch). We leverage the Malai architecture to model the system under test to establish two contributions: the definition of novel adequacy criteria to generate test cases that cover Malai models; an algorithm for the automatic generation of test suites that satisfy the adequacy criteria. We applied the novel approach to two open-source interactive systems to validate the ability of generated test cases to reveal bugs. This early work is part of the project Connexion (*cf.* Section 8.1.3) which notably focuses on testing interactive parts of critical systems.

URBANET Team

6. New Results

6.1. Scalable protocols for capillary networks.

Participants: Ibrahim Amadou, Quentin Lampin, Bilel Romdhani, Alexandre Mouradian, Isabelle Augé-Blum, Fabrice Valois

6.1.1. Beacon-less and opportunistic routing.

During the thesis of Ibrahim Amadou [1], we were focused on the issues of energy in WSNs through energy-efficient routing and medium access control protocols. The contributions of research work can be summarized as follows. First, we were interested on the energy issues at the routing layer for multi-hop wireless sensor networks (WSNs). We proposed a mathematical framework to model and analyze the energy consumption of routing protocols in multi-hop WSNs by taking into account the protocol parameters, the traffic pattern and the network characteristics defined by the medium channel properties, the dynamic topology behavior, the network diameter and the node density. We showed that Beacon-less routing protocol is a good candidate for energy saving in WSNs.

We investigated the performance of some existing relay selection schemes which are used by Beacon-less routing protocols. Extensive simulations were realized in order to evaluate their performance locally in terms of packet delivery ratio, duplicated packet and delay. Then, we extended the work in multi-hop wireless networks and developed an optimal solution, Enhanced Nearest Forwarding within Radius, which tries to minimize the per-hop expected number of retransmissions in order to save energy.

We presented a new Beacon-less routing protocol called Pizza-Forwarding (PF) without any assumption on the radio environment: neither the radio range nor symmetric radio links nor radio properties (shadowing, etc.) are assumed or restricted. A classical greedy mode is proposed. To overcome the hole problem, packets are forwarded to an optimal node in the two hop neighbor following a reactive and optimized neighborhood discovery.

In order to save energy due to idle listening and overhearing, we proposed to combine PF's main concepts with an energy-efficient MAC protocol to provide a joint MAC/routing protocol suitable for a real radio environment. Performance results lead to conclude to the powerful behavior of PF-MAC.

In collaboration with Orange Labs, we designed QOR, an opportunistic routing protocol for wireless sensor networks [16]. QOR first builds a stable directed acyclic logical routing structure and a prefix-based addressing plan stemming from data sinks. This addressing plan is then used to define the potential forwarders set for each source and allows a strict scheduling and an unique selection of the forwarder for each transmission thanks to a cascading acknowledgment scheme. QOR is particularly suited for sensor networks that require high delivery ratio under severe energy constraints. Extensive simulations show the benefits of QOR over an implementation of the IETF routing protocol for Lossy and Low Power networks, RPL, tailored to provide high delivery ratios. Our case studies shows that QOR saves up to 50% energy and reduces the end-to-end delay of a factor of 4 times while maintaining similar delivery ratios.

Most existing routing protocols designed for WSNs assume that links are symmetric, which is in contradiction with what is observed in the field. Indeed, many links in real-world WSNs are asymmetric. Asymmetric links can dramatically decrease the performance of routing algorithms not designed to cope with them. Quite naturally, most existing routing protocol implementations prune the asymmetric links to only use the symmetric ones. In our experience, asymmetric links are a valuable asset to improve network connectivity, capacity and overall performance [20],[2]. We therefore introduced AsymRP (Asymmetric Convergecast Routing Protocol) [21], a new routing protocol for collecting data in WSNs. AsymRP assumes 2-hop neighborhood knowledge and uses implicit and explicit acknowledgment. It takes advantage of asymmetric links to increase delivery ratio while lowering hop count and packet replication.

6.1.2. MAC and cross-layer mechanisms for QoS.

Protocols developed during the last years for Wireless Sensor Networks (WSNs) are mainly focused on energy-consumption optimization and autonomous mechanisms (e.g. self-organization, self-configuration, etc). Nevertheless, with new WSN applications appear new QoS requirements such as time constraints. Real-time applications require the packets to be delivered before a known time bound which depends on the application requirements. We particularly focused on applications which consist in alarms that are sent to the sink node (e.g. air pollution monitoring). We proposed the Real-Time X-layer Protocol (RTXP) [27], a real-time communication protocol that integrates mechanisms for both MAC and routing layers. Our proposal aims at guaranteeing an end-to-end constraint delay, while keeping good performances on other parameters, such as energy consumption. For this purpose the protocol relies on a hop-count-based Virtual Coordinate System (VCS) which classifies nodes having the same hop-count from the sink, allows forwarder selection, and gives to the nodes an unique identifier in a 2-hop neighborhood allowing deterministic medium access. This protocol has better performances than state-of-the-art protocols, in terms of time constraints and reliability, even with unreliable radio links.

In the ARESA2 project, but also in a joint collaboration with Orange Labs, we studied receiver initiated MAC protocol to compare their performance to the more classical receiver-based MAC one [17]. We proposed the Self Adapting Receiver Initiated MAC protocol (SARI-MAC), a novel asynchronous MAC protocol for energy constrained Wireless Sensor Networks. SARI-MAC self-adapts to the traffic load to meet specified Quality of Service requirements at the lowest energy cost possible. To do so, SARI-MAC relies on traffic estimation, duty-cycle adaptation and acknowledgment mechanisms. Our performance evaluation assesses that SARI-MAC meets given QoS requirements in a energy efficient manner and outperforms the state of the art protocol RI-MAC in a broad range of traffic scenarios.

For energy constrained wireless sensor networks, lifetime is a critical issue. Several medium access control protocols have been proposed to address this issue, often at the cost of poor network capacity. To address both capacity and energy issues, we proposed a novel medium sharing protocol for Wireless Sensor Networks named Cascading Tournament (CT-MAC) [15]. CT-MAC is a synchronous, localized, dynamic, joint contention/allocation protocol. Relying on cascading iterations of tournaments, CT-MAC allocates multiple time slots to nodes that compete for accessing the medium. CT-MAC offers an unprecedented trade-off between traffic delay, network capacity and energy efficiency and stands out as a solid candidate for energy constrained sensor networks that must support heterogeneous traffic loads. Our simulations show that CT-MAC significantly outperforms the state-of-the-art SCP- MAC protocol.

6.2. Characterizing urban capillary wireless networks.

Participants: Sandesh Uppoor, Diala Naboulsi, Rodrigue Domga Komguem, Anis Ouni, Alexandre Mouradian, Isabelle Augé-Blum, Hervé Rivano, Marco Fiore, Fabrice Valois

6.2.1. Properties of urban road traffic of interest to mobile networking.

The management of mobility is commonly regarded as one of the most critical issues in large-scale telecommunication networks. The problem is exacerbated when considering vehicular mobility, which is characterized by road-constrained movements, high speeds, sudden changes of movement direction and acceleration, and significant variations of these dynamics over daytime. The understanding of the properties of car movement patterns becomes then paramount to the design and evaluation of network solutions aimed at vehicular environments.

We first analyzed how the vehicular mobility in a large-scale urban region affects a cellular infrastructure intended to support on-board users. We studied the spatial and temporal distribution of traffic load induced by vehicular users, their spatial flows, their inter-arrival and residence times at cells [22].

We then studied the topological features of a network built on moving vehicles, considering the instantaneous connectivity of the system [28]. Our results evidence the spatial and temporal diversity of road traffic, stressing the importance of a correct modeling of road traffic towards the reliable performance evaluation of network protocols. Additionally, the results outline how commonly adopted assumptions (e.g., Poisson user arrivals at the network base stations) do not hold under vehicular environments, and how the V2V-based network has low connectivity, availability, reliability and navigability properties.

6.2.2. The limits of RSSI-based localization.

Numerous localization protocols in Wireless Sensor Networks are based on Received Signal Strength Indicator. Because absolute positioning is not always available, localization based on RSSI is popular. More, no extra hardware is needed unlike solutions based on infra-red or ultrasonic. Moreover, the theory gives a RSSI as a function of distance. However, using RSSI as a distance metric involves errors in the measured values, resulting path-loss, fading, and shadowing effects. We did experimentation results from three large WSNs, each with up to 250 nodes [23]. Based on our findings from the 3 systems, the relation between RSSI and distance is investigated according to the topology properties and the radio environment. We underline the intrinsic limitations of RSSI as a distance metric, in terms of accuracy and stability. Contrary to what we assumed, collaborative localization protocol based on Spring-Relaxation algorithm can not smooth the distance-estimation errors obtained with RSSI measurements.

6.2.3. Modeling and optimization of wireless networks.

In critical real-time applications, when an event is detected, the Worst Case Traversal Time (WCTT) of the message must be bounded. However, despite this, real-time protocols for WSNs are rarely formally verified. The model checking of WSNs is a challenging problem for several reasons. First, WSNs are usually large scale so it induces state space explosion during the verification. Moreover, wireless communications produce a local broadcast behavior which means that a packet is received only by nodes which are in the radio range of the sender. Finally, the radio link is probabilistic. The modeling of those aspects of the wireless link in model checking is not straightforward and it has to be done in a way that mitigates the state space explosion problem. We are currently working on proposing a methodology adapted to WSNs, and based on Timed Automata (TA) and model-checking. First results are promising [19], but needed to be further investigated.

While the large variety of routing protocols (geographical, gradient, reactive, ...) proposed in the literature provide a set of pertinent solutions for optimizing the energy consumption for multi-hop wireless networks, they do not permit to know the conditions of use of these protocols based on parameters such as: the dynamics of topology, traffic pattern, the density and diameter of the network, the load, etc. In [12], we presented a theoretical model for evaluating the energy consumption for communication protocols taking into account both the dynamics of nodes and links, the properties of topology, the traffic pattern, the control/data packets and a realistic channel model. This model is applied successively to several protocols (GPSR, AODV, OLSR and PF) to highlight their optimum usage and it permits to conclude that Beacon-Less routing protocols are adapted for application with low traffic.

We continued developing optimization tools for building optimal solution to various problems of multi-hop wireless networks. Most of these contributions combine graph theoretical basis with Mixed Integer Linear Programming techniques, and are valuable for understanding the extremal behaviors of the systems and guide the development of efficient architectures and protocols. In this sense, we have considered a new edge coloring problem to model call scheduling optimization issues in wireless mesh networks: the proportional coloring [6]. It consists in finding a minimum cost edge coloring of a graph which preserves the proportion given by the weights associated to each of its edges. We show that deciding if a weighted graph admits a proportional coloring is pseudo-polynomial while determining its proportional chromatic index is NP-hard. We then give lower and upper bounds for this parameter that can be computed in pseudo-polynomial time. We finally identify a class of graphs and a class of weighted graphs for which the proportional chromatic index can be exactly determined.

Dealing with wireless mesh network, we have investigated the fundamental trade-off between transmitting energy consumption and network capacity [24]. The results on this trade-off have been computed using MILP models solved with column generation techniques. The main contribution relies in the ability to consider a realistic SINR model of the physical layer with a continuous power control and discrete transmission rate selection at each node. In order to model these functionalities, a strong formulation (in the sense that the linear relaxation gives relevant lower bounds) of the rate selection is introduced.

The behavior of beaconless geographic forwarding protocols for wireless sensor networks has also been modeled [9]. A realistic physical layer is taken into account by combining MILP models with simulation based inputs on the number of required retransmissions for realizing a transmission. The model is then able to compute energy efficient routings and allows for understanding the most efficient relay selection schemes, denoted Furthest Forward within Reliable neighbors (FFRe).

6.3. Solutions for cellular networks.

Participants: Anis Ouni, Fabrice Valois, Hervé Rivano, Marco Fiore

6.3.1. Content downloading through a vehicular network.

We considered a system that leverages vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communication to transfer large contents to users on-board moving cars. This paradigm is intended to relieve the cellular infrastructure from the high load that such downloads would induce, once vehicles are widely equipped with infotainment devices.

We first characterized the theoretical performance limits of such a vehicular content downloading system by modeling the downloading process as an optimization problem, and maximizing the overall system throughput. Our approach allows us to investigate the impact of different factors, such as the roadside infrastructure deployment, the vehicle-to-vehicle relaying, and the penetration rate of the communication technology, even in presence of large instances of the problem [7]. We then evaluated practical protocols for vehicular downloading, devising solutions for the selection of relay vehicles and data chunks at the Road Side Units (RSUs), and evaluating them in real-world road topologies, under different infrastructure deployment strategies [8].

Our results show that V2V transfers can significantly increase the download rate of vehicular users in urban/suburban environments, and that such a result holds throughout diverse mobility scenarios, RSU placements and network loads. Also, they highlight the existence of two operational regimes at different penetration rates and the importance of an efficient, yet 2-hop constrained, V2V relaying.

6.3.2. Toward green mesh and cellular networks.

On the one hand, a promising technique for minimizing the transmission power of cellular networks seems to be a dramatic densification of micro-cells coverage. On the other hand, increasing the number of micro-cells multiplies the energy consumed by the cells whatever their state, idle, transmitting or receiving. For a sustainable deployment of such micro-cell infrastructures and for a significant decrease of the overall energy consumption, an operator needs to be able to switch off cells when there are not absolutely needed. The densification of the cells induces the need for an autonomic control of the on/off state of cells. This has motivated a preliminary investigation on exploiting within the micro-cellular settings the manifold results of duty cycles for Wireless Sensor Networks where switching nodes on and off is done in a distributed or localized manner while coverage and connectivity properties are maintained [29].

Focusing on broadband wireless mesh networks based on OFDMA resource management, and considering a realistic SINR model of the physical layer with a continuous power control and discrete transmission rate selection at each node, we have investigated the trade-off between transmission energy consumption and network capacity [24]. Correlation between capacity and energy consumption is analyzed as well as the impact of physical layer parameters - SINR threshold and path-loss exponent. We highlight that there is no significant tradeoff between capacity and energy when the power consumption of idle nodes is important. We also show that both energy consumption and network capacity are very sensitive to the SINR threshold variation. We also highlight that power control and rate selection are not expandable to an optimal system configuration.

6.4. Miscellaneous security issues in capillary networks.

Participants: Ochirkhand Erdene-Ochir, Fabrice Valois, Marco Fiore

6.4.1. Resiliency in routing protocols.

Within the ARESA2 project, we defined the notion of resiliency for routing protocols in wireless sensor networks and we applied it to several routing strategies to provide an understandable taxonomy [3]. Efforts have been made to compare routing protocols according to their resiliency in wireless multi-hop sensor networks in the presence of packet dropping malicious insiders. In [13], we proposed a new taxonomy of routing protocols obtained by applying our resiliency metric. Several resiliency enhancing methods such as introducing a random behavior to the classical routing protocols and a new data replication method based on the distance information have been evaluated as well. Simulation results demonstrate the effectiveness of the proposed approach.

6.4.2. Verifying the positions announced by mobile nodes.

A growing number of ad hoc networking protocols and location-aware services require that mobile nodes learn the position of their neighbors. However, such a process can be easily abused or disrupted by adversarial nodes. In absence of a-priori trusted nodes, the discovery and verification of neighbor positions presents challenges that have been scarcely investigated in the literature.

We proposed a fully-distributed cooperative solution that is robust against independent and colluding adversaries. Results show that our protocol can thwart more than 99% of the attacks under the best possible conditions for the adversaries, with minimal false positive rates [5].

A centralized solution was also developed, that leverages anonymous position beacons from vehicles, and the cooperation of nearby cars collecting and reporting the beacons they hear. Such information allows an authority to verify the locations announced by vehicles, or to infer the actual ones if needed [18].

ALICE Project-Team

5. New Results

5.1. A Runtime Cache for Interactive Procedural Modeling

Participant: Sylvain Lefebvre.

This work further explores hashing techniques that we developed over the past years. In particular, we considered modifying our hashing scheme to create a run-time cache. The cache avoids expensive computations when texturing implicit surfaces with complex procedural functions. This is a result from a collaboration with the Karlsruhe Institute of Technology which was funded by an Inria COLOR grant and has been published this year in the journal "Computers & Graphics" [14].

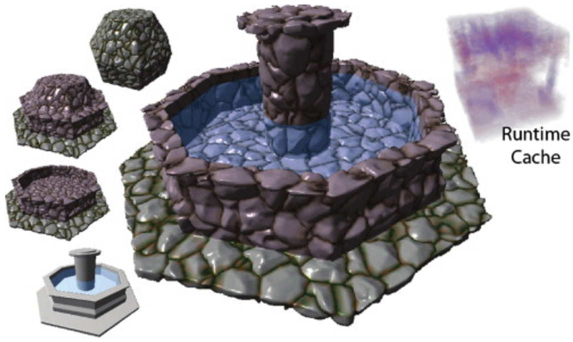


Figure 1. A Runtime Cache for Interactive Procedural Modeling.

5.2. Texture Synthesis

Participants: Sylvain Lefebvre, Bruno Jobard.

We continued investigating on Gabor Noise and considered fitting the parameters of our Gabor noise texturing technique from example images. This required a new formulation of our noise, allowing us to solve the problem as a basis pursuit denoising optimization. This is the result of a collaboration with the team REVES / Inria Sophia-Antipolis, the Katholieke Universiteit of Leuven and Université Paris Descartes. This work has been presented at the SIGGRAPH conference this year [8].

We also revisited techniques for texture synthesis explicitly copying and assembling large patches of an example image to form a new texture. We accelerate this process through a parallel implementation which both optimizes for the shape of the patches and a deformation along their boundary to better match edges. This work is part of the PhD thesis of Anass Lasram and has been presented this year at the Eurographics/ ACM SIGGRAPH Symposium on High Performance Graphics, [19].

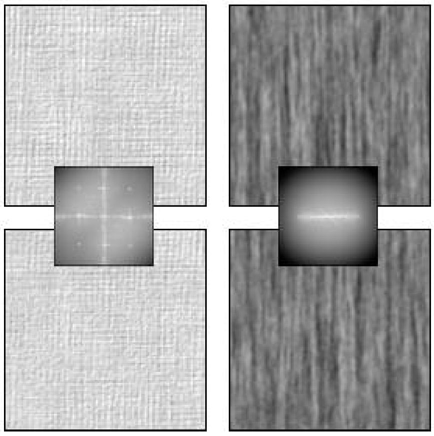


Figure 2. Gabor Noise by Example.

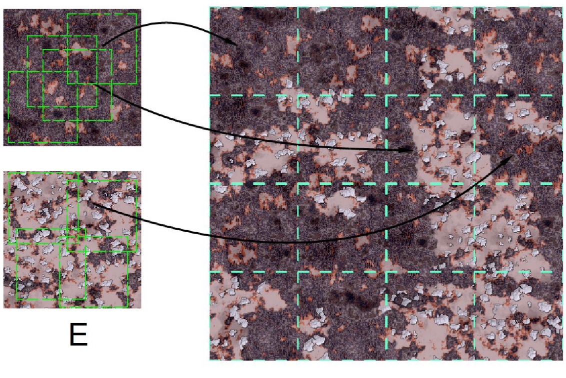


Figure 3. Parallel patch–based texture synthesis.

We also studied ways of helping the user to select the parameters of procedural texture generators, by proposing two contributions :

- We studied how to summarize the appearances generated by complex procedural textures in a small preview image. The challenge is to capture the large variety of appearances despite a limited pixel space. We formulate the problem as a layout of high-dimensional samples in a regular grid, and optimize for it through a modified Self Organizing Map algorithm. This work is part of the PhD thesis of Anass Lasram, and is a collaboration with our industrial partner Allegorithmic. This work has been published this year in the journal "Computer Graphics Forum", [10].
- The parameters of complex procedural textures are typically chosen through a slider-based interface. We augment this interface with preview images which predict how the texture will change when manipulating the slider. This greatly simplifies the process of choosing parameters for these textures. This work is part of the PhD thesis of Anass Lasram, and is a collaboration with our industrial partner Allegorithmic. This work has been published this year as EUROGRAPHICS short paper, [18].

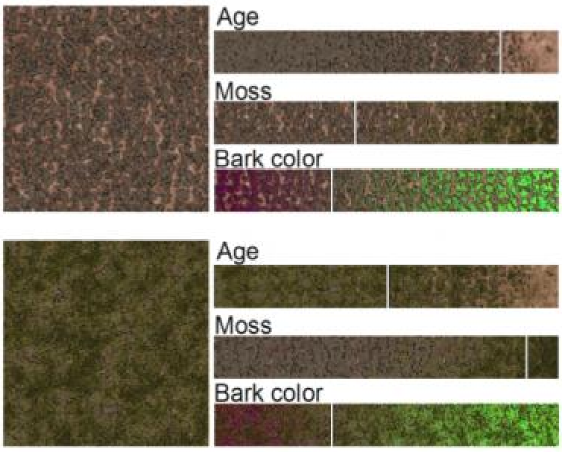


Figure 4. Scented Sliders for Procedural Textures.

5.3. Algorithms and analysis

Participants: Laurent Alonso, Samuel Hornus.

Data structure for fast witness complexes: Samuel Hornus is currently pursuing work started while a post-doc in Sophia Antipolis, on data structure for the fast construction of witness complexes; these are sub complexes of Delaunay triangulations that can be faster to compute for low dimensional data embedded in high dimensional ambiant space.

Analysis of Boyer and Moore's MJRTY Algorithm: Given a set of n elements each of which is either red or blue, Boyer and Moore's algorithm uses pairwise equal/not equal color comparisons to determine the majority color. We analyze the average behavior of their algorithm, proving that if all 2^n possible inputs are equally likely, the average number of color comparisons used is $n - \sqrt{2n/\pi} + O(1)$ and have variance in $\frac{\pi-2}{\pi}n - \frac{\sqrt{2\,n}}{\sqrt{\pi}} + O(1)$. This work has been submitted to SIAM Journal On Computing.

5.4. Visualizing $2D$ Flows with Animated Arrow Plots

Participants: Bruno Jobard, Nicolas Ray, Dmitry Sokolov.

Flow fields are often represented as a set of static arrows in illustration of scientific vulgarization, documentary, meteorology, etc. This simple and schematic representation lets an observer intuitively interpret the main properties of a flow: its orientation and velocity magnitude (Figure 5).

We have investigated how to automatically generate dynamic versions of such representations for $2D$ unsteady flow fields. As a result, we designed an algorithm able to smoothly animate arrows along the flow while controlling their density in the domain over time. Beside keeping an even distribution of arrows over time, we made significant efforts to remove disturbing rendering artefacts such as the apparition of a new arrow, the removing of existing arrows, and the representation of field where the velocity is null. This work has been published as a research report, [24].

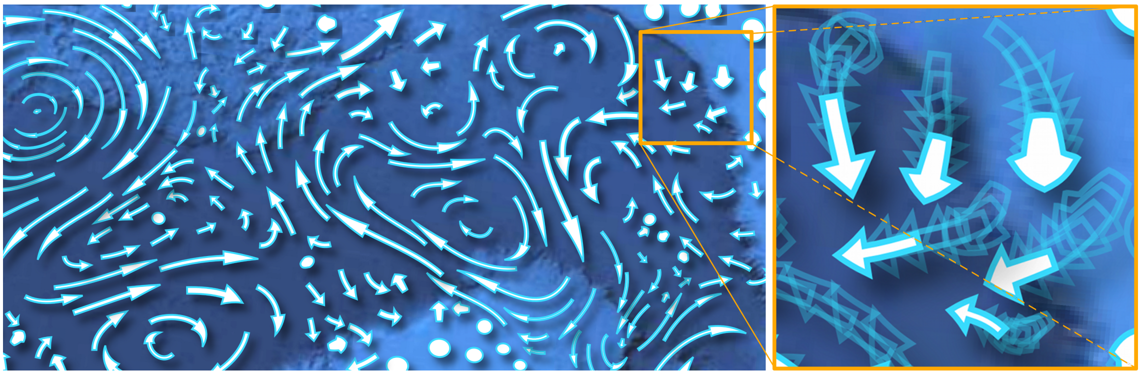


Figure 5. Ocean currents visualized with a set of dynamic arrows. The Close-up shows the arrow trajectories and the morphing of their glyphs.

5.5. Fixing normal constraints for generation of polycubes

Participants: Nicolas Ray, Dmitry Sokolov.

A polycube is a piecewise linearly defined surface where all faces are squares that are perpendicular to an axis of a global basis. Deforming triangulated surfaces to polycubes provides maps (form the original surface to the polycube) that can be used for a number of applications including hex-meshing. To define such a deformation, it is necessary to determine, for each point of the original surface, what will be its orientation (global axis) in the polycube.

This problem is actually tackled by heuristics that basically affect the closest global axis to the surface normal. Coupled with an mesh deformation as pre-processing and some fixing rules as a post-processing, it is able to provide nice results for a number of surfaces. However, nothing ensures that the surface can be deformed to a polycube having these desired face orientation.

We have worked on a method able to determine if there exists a deformation of the surface that respects a given orientation constraint on each point. We have also design an automatic solution that can fix constraints that would prevent the existence of a deformation into a polycube (Figure 6).

This study has highlighted that the constraints on desired orientation are global and requires constrained optimization methods to be solved. Our current solution is able to manage many cases where previous works would fail, but we can still produce some complex cases where interactions between dimension may lead to deadlocks.

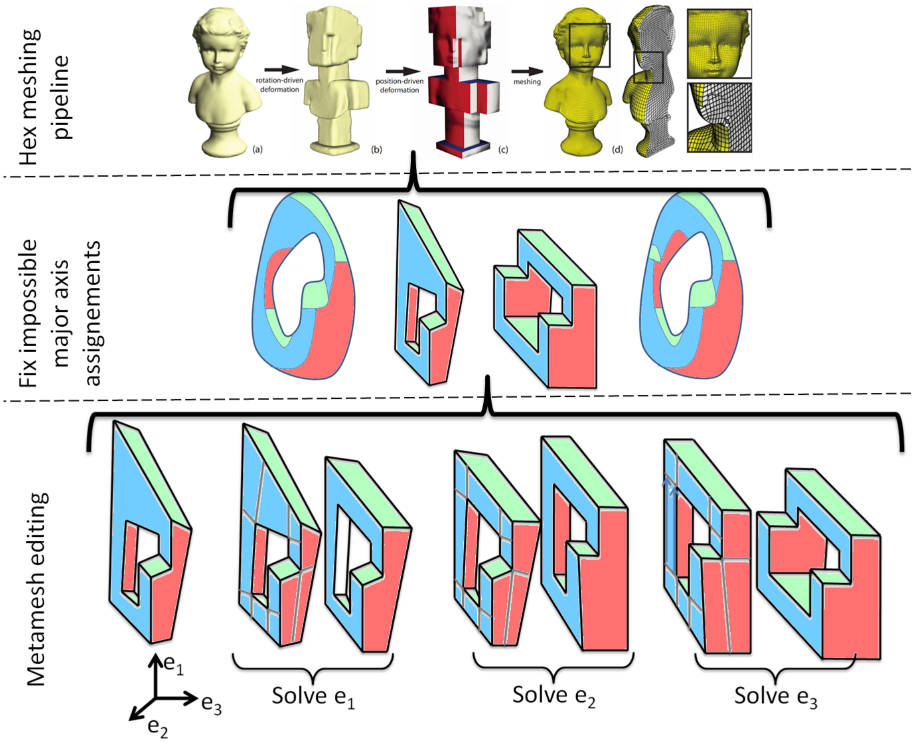


*Figure 6. **Upper row:** the surface is deformed to make its normals closer to major axis, but to reach an equality, we need to have a coherent "wished orientation" of the faces. **Middle row:** we define a valid deformation into a polycube by editing the "wished orientation". **Lower row** the resolution is performed a dimension at a time.*

5.6. Control of the differential behaviour of the joining curve between two fractal curves

Participants: Dmitry Sokolov, S. Podkorytov, C. Gentil, S. Lanquetin.

The general objective of our work is to create a geometric modeller based on iterative processes. Iterative processes can be used to describe a wide array of shapes inaccessible to standard methods such as fractal curves or sets. Our work is based on Boundary Controlled Iterative System (BCIFS). BCIFS upgrades the standard iterative process such as Iterated Function System (IFS) with B-Rep structure. We can describe objects with familiar B-rep structure, where each cell is a fractal object. For instance, if we consider a polyhedron, then each face is a fractal surface, and each edge is a fractal curve. Objects modelled with BCIFS not necessary have the fractal properties, objects such as B-splines curves and surfaces can be modelled as well. So with BCIFS formalism we can operate with both standard and fractal objects.

With this objective in mind, we have to provide tools that work with fractal objects in the same manner as with objects of classical topology. In this project we focus on the constructing of an intermediate curve between two other curves defined by different iterative construction processes. Similar problem often arises with subdivision surfaces, when the goal is to connect two surfaces with different subdivision masks. We start by dealing with curves, willing to later generalize our approach to surfaces. We formalise the problem with Boundary Controlled Iterated Function System model. Then we deduct the conditions that guaranties continuity of the intermediate curve. These conditions determine the structure of subdivision matrices. By studying the eigenvalues of the subdivision operators, we characterise the differential behaviour at the connection points between the curves and the intermediate one. This behaviour depends on the nature of the initial curves and coefficients of the subdivision matrices. We also suggest a method to control the differential behaviour by adding intermediate control points (Figure 7). This work was presented at the Symposium on Solid and Physical Modeling [23].

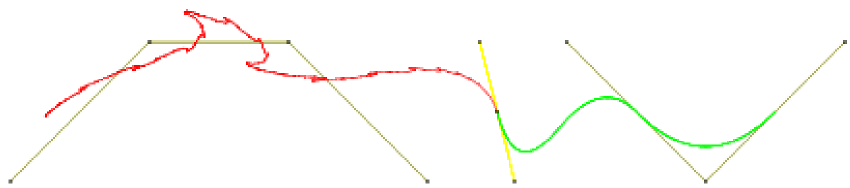


Figure 7. Two intermediate curves between the fractal curve and B-spline. Three control point are used to control the shape of the curve

5.7. Approximate convex hull of affine iterated function system attractors

Participants: Dmitry Sokolov, A. Mishkinis, C. Gentil, S. Lanquetin.

In this paper, we present an algorithm to construct an approximate convex hull of the attractors of an affine iterated function system (IFS). We construct a sequence of convex hull approximations for any required precision using the self-similarity property of the attractor in order to optimize calculations. Due to the affine properties of IFS transformations, the number of points considered in the construction is reduced. The time complexity of our algorithm is a *linear* function of the number of iterations and the number of points in the output convex hull. The number of iterations and the execution time increases logarithmically with increasing accuracy. In addition, we introduce a method to simplify the approximation of the convex hull without loss of accuracy. Figure 8 gives and illustration. This work was published at the Chaos, Solitons & Fractals journal [12].

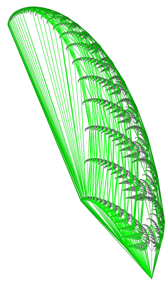


Figure 8. Approximate convex hull for a 3D IFS attractor.

5.8. Shift-Based Parallel Image Compositing on InfiniBand Fat-Trees

Participant: Xavier Cavin.

In this work, we propose a new parallel image compositing algorithm, called Shift-Based, relying on a well-known communication pattern called shift permutation. Indeed, shift permutation is one of the possible ways to get the maximum cross bisectional bandwidth provided by an InfiniBand fat-tree cluster. We show that our Shift-Based algorithm scales on any number of processing nodes (with peak performance on specific counts), allows overlapping communications with computations and exhibits contention free network communications. This is demonstrated with the image compositing of very high resolution images at interactive frame rates. This work is a collaboration with the SED service of Inria (Olivier Demengeon). It has been presented this year at the Eurographics Symposium on Parallel Graphics and Visualization, [17].

5.9. Multi view data processing

Participants: Rhaleb Zayer, Alejandro Galindo, Kun Liu.

Direct use of denoising and mesh reconstruction algorithms on point clouds originating from multi-view images is often oblivious to the reprojection error. This can be a severe limitation in applications which require accurate point tracking, e.g., metrology. we propose a method for improving the quality of such data without forfeiting the original matches. We formulate the problem as a robust smoothness cost function constrained by a bounded reprojection error. The arising optimization problem is addressed as a sequence of unconstrained optimization problems by virtue of the barrier method. Experimental results on synthetic and acquired data compare our approach to alternative techniques. This work has been presented this year at the 8th International Symposium on Visual Computing, [20].

5.10. Deformation modeling of slender objects

Participants: Rhaleb Zayer, Alejandro Galindo, Kun Liu.

A desirable property when modeling/editing slender curve-like objects is the ability to emulate the deformation behavior of natural objects (e.g. cables, ropes). Taking such physical considerations into account needs also to abide to editing requirements such as interactivity and full access and control of all degrees of freedom (positional and rotational constraints) during interaction. We regard editing as a static deformation problem but our treatment differs from standard finite element methods in the sense that the interpolation is based on deformation modes rather than the classic shape functions. A careful choice of these modes allows capturing the deformation behavior of the individual curve segments, and devising the underlying mathematical model from simple and tractable physical considerations. In order to correctly handle arbitrary user input (e.g.

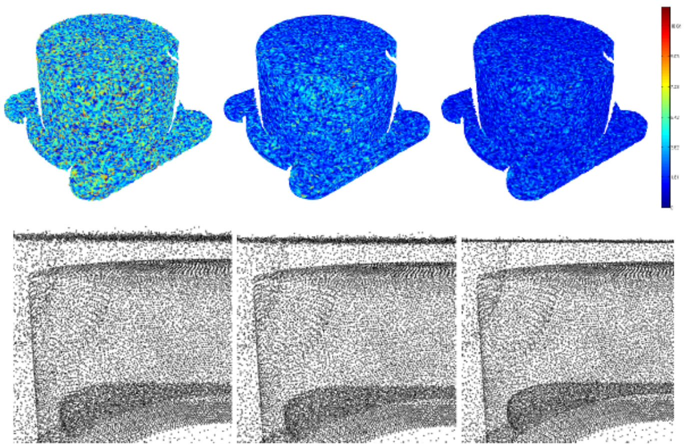


Figure 9. Example of denoising.

dragging vertices in a fast and excessive manner), our approach operates in the nonlinear regime. The arising geometric nonlinearities are addressed effectively through the modal representation without requiring complicated fitting strategies. In this way, we circumvent commonly encountered locking and stability issues while conveying a natural sense of flexibility of the shape at hand. Experiments on various editing scenarios including closed and non-smooth curves demonstrate the robustness of the proposed approach. This work has been published this year in the journal "Computers & Graphics", [15].

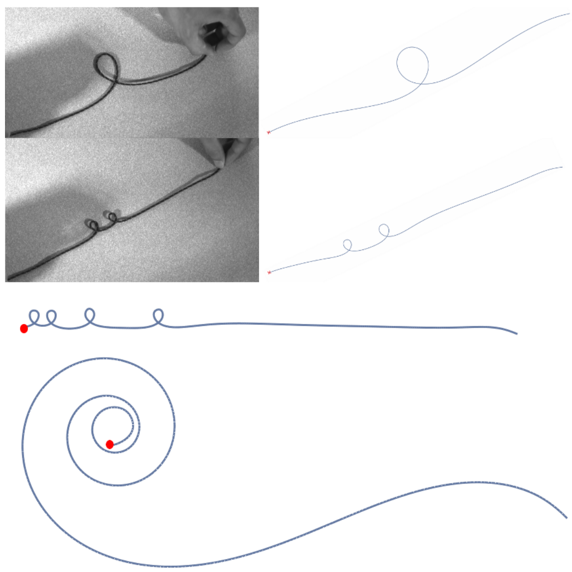


Figure 10. Example of curves.

5.11. Temporally consistent 3D meshing from video data

Participants: Dobrina Boltcheva, Phuong Ho, Bruno Lévy.

This work is a part of the ANR Morpho project (Morpho) which aims at combined analysis of human shapes and motions. In particular, the goal is to study how motions relate to human shapes or how shapes deform in typical motions. During this year, we addressed the first challenge which is building temporally consistent 3D meshes from silhouette images. We have already achieved a very fast meshing algorithm for each frame based on the Centroidal Voronoi Tessellation which has been previously developed in our team. Actually, we are investigating different ways for adding the temporal consistency within our optimisation framework. In particular, we are studding a strategy based on the optimal transport paradigm.

5.12. Re-meshing surfaces

Participants: Nicolas Bonneel, Bruno Lévy, David Lopez, Vincent Nivoliers, DongMing Yan.

In the frame of the ERC GOODSHAPE project, we continued to develop new methods to optimize the sampling of 3D objects. In particular, we studied how to sample a surface with generalized primitives, such as line segments and deformable graphs [11]. We also focused on the problem of remeshing a surface with quads, or fiting a polynomial surface to an input mesh. We proposed a method that minimizes an approximation of the integrated squared distance, based on a restricted Voronoi diagram [22]. Still on the same topic of mesh quadrangulation, we co-published a survey with other international experts of this field [16].

We also worked on anisotropic surface meshing, and developed a technique based on embedding into higher dimensional space and a fast computation of the restricted Voronoi diagram [21].

ALPAGE Project-Team

6. New Results

6.1. Advances in symbolic and hybrid parsing with DyALog and FRMG

Participants: Éric Villemonte de La Clergerie, François Barthélemy, Julien Martin.

Within the team is developed a wide-coverage French meta-grammar (FRMG) and a efficient hybrid TAG/TIG parser based on the DYALOG logic programming environment [120] and on the Le*fff* morphological and syntactic lexicon [105]. It relies on the notion of factorized grammar, themselves generated from a representation that lies at a higher level of abstraction, named Meta-Grammars [122]. At that level, linguistic generalizations can be expressed, which in turn makes it possible to transfer meta-grammars from one language to a closely related one. The hybrid TAG/TIG parser generator itself implements all kinds of parsing optimizations: lexicalization (in particular via hypertags), left-corner guiding, top/bottom feature analysis, TIG analysis (with multiple adjoining), and others. The recent evolutions go towards an hybridization with statistical approaches.

6.1.1. Tuning FRMG's disambiguation mechanism

Continuing works initiated in 2011 on the exploitation of the dependency version of the French TreeBank (FTB), Éric de La Clergerie has explored the tuning of FRMG's rule base disambiguation mechanism using a larger set of features and weight learned from the FTB. In 2011, this approach led to on improvement from 82.31% to 84.54% in terms of accuracy (LAS - Labelled Attachment Score) on the test part of the FTB. By increasing the set of features, in particularly using higher-order dependency features (on parent edge and sibling edges), and a better understanding of the iterative tuning mechanism, it was possible to reach 85.95% LAS. This tuning mechanism is based on the idea of adding or subtracting some weight to a disambiguation rule given some specific contexts (provided by the features), where the delta is progressively learned from the accuracy of the disambiguation rule in terms of edge selection or rejection. The learning algorithm presents some relationships with the perceptron approach, but the use of a more standard implementation of the perceptron led to less interesting gains.

During the same time, the coverage of FRMG was improved (to reach for instance 94% of full parses on the FTB).

6.1.2. Synchronous Tree-Adjoining Grammars

A preliminary work has been done to implement *Synchronous Tree-Adjoining Grammars* (STAGs) in DYALOG, relying on the notion of *Thread Automata* [119]. Synchronous Tree Adjoining Grammars is an instance of formalism where the order of the components of a tree structure is not fully determined. This leads to combinatorial alternatives when parsing, while a tree-structure corresponding to the input string has to be build. A specific front-end has been written to implement STAGs. The work on the back-end is still in progress, with the goal to have a common intermediate representation for several mildly context-sensitive formalisms where some node operations non-deterministically pick a node out of a finite set of nodes. STAGs are an instance of such formalisms, Multi-Component Tree Adjoining Grammars (MCTAGs) are another instance. The intermediate representation consists in Thread Automata (TA), an extension of Push-Down Automata where several threads of computations are considered and only one is active at any time.

6.1.3. Adding weights and probabilities to DyALog

Weights can already be used during the disambiguation phase of the FRMG parser, implemented in DYALOG. However, a deeper implementation of weights and probabilities in DYALOG was initiated in 2012 by Julien Martin during his Master internship. By enriching the structure of the backpointers (relating the items to their parent items), it is now possible to maintain an ordered weighted list of derivations, to update the scheduling of items wrt their weight, to update the weights of all the descendants of an item I when updating I's weight. The motivation is of course to be able to favor the best analysis first during parsing. A second objective (which has been implemented) is the possibility to extract the n-best parses after parsing (but keeping a shared derivation forest). A third objective, remaining to be done, is related to the use of beam search techniques to prune the search space during parsing. A longer-term objective is the abstraction of this work to be able to work on semi-rings.

6.2. Tree transformation

Participants: Éric Villemonte de La Clergerie, Corentin Ribeyre, Djamé Seddah.

In 2011, the conversion of native FRMG dependencies into the CONLL dependency scheme was the occasion to explore new ideas about tree transformation (for dependencies), based on the notion of two-level transformation with a first level relying on local transformation rules and a second level being controlled by constraints carried by the first level edges. During his Master internship, Corentin Ribeyre has formalized and re-implemented this approach in a more systematic and generic way. This work was also completed by the use of example-based learning techniques to quickly learn the local transformation rules of the first level. The line of research is motivated by possibility to quickly develop a reduced set of transformation rules (thanks to the examples and the constraint level) for a large variety of applications, such as information extration but also conversion toward a deep syntax level or a shallow semantic level. A poster paper was presented at TAG+11 [29].

6.3. lexical knowledge acquisition and visualization

Participants: Éric Villemonte de La Clergerie, Mickael Morardo, Benoît Sagot.

In relation with our collaboration with Lingua & Machina (cf section 4.4), Mikael Morardo has enriched the interfaces of the WEB platform Libellex for the visualization and validation of more complex lexical resource. In particular, the focus has been on the development of a graph-based view with the javascript Library `d3.js` to represent large lexical networks. The current implementation is powerful enough to deal with large networks of several teens of thousands of connections, allowing the visualization of fragments of the network and an easy navigation. Because the graph-view proved to be both intuitive and efficient, the previous list-based view for terminology was partially re-implemented in the new graph-view. It was also extended for visualizing and validating more complex lexical networks, like the French Wordnet WOLF coupled with the original English WordNet (cf 5.9).

The graph-based view was used to explore several networks built using Harris' distributional hypothesis (through a clustering algorithm) on the output of FRMG for several corpora. Because terminology was now be visualized at the same time, the clustering algorithm was modified to be able to take into account a list of terms (also automatically extracted from the parsed corpora) .

6.4. Advances in statistical parsing

Participants: Marie Candito, Benoît Crabbé, Djamé Seddah, Enrique Henestroza Anguiano.

6.4.1. Statistical Parsing

We have achieved **state-of-the art results for French statistical parsing**, adapting existing techniques for French, a language with a morphology richer than English, either for constituency parsing [110], [113] or dependency parsing [68]. We made available The Bonsai parsing chain [1] (cf. 5.4), that gathers preprocessing tools and models for French dependency parsing into an easy-to-use parsing tool for French. We designed our parsing pipeline with modularity in mind: our parsing models are interchangeable. For instance, dependencies output can either be generated from a PCFG-LA based parser associated with a functional role labeler or from any dependency parsers trained on our dependency treebank [68]. Tokens can either be raw words, POS tagged lemmas or word clusters [69].

[1] http://alpage.inria.fr/statgram/frdep/fr_stat_dep_parsing.html

We have innovated in the tuning of tagsets to optimize both grammar induction and unknown word handling [75], thus providing the best parsing models for French [111]. Then we have contributed on three main points:

1. conversion of the French Treebank [55] used as constituency training data into a dependency treebank [4], which is now used by several teams for dependency parsing;
2. an original method to reduce lexical data sparseness by replacing tokens by unsupervised word clusters, or morphological clusters [64], [112];
3. a postprocessing step that uses specialized statistical models for parse correction [81].

For the last 18 to 12 months, we have been increasingly focused in increasing the robustness of our parsing models by (a) validating our approach on other morphologically-rich languages; (b) other domains and (c) on user generated content. All of those challenging the current state-of-the-art in statistical parsing.

6.4.2. Multilingual parsing

Applying the techniques we developed for reducing lexical data, which is commonly found in morphologically-rich languages (MRLs) and optimizing the POS tagset, we integrated lexical information through data driven lemmatisation [112] and POS tagging [79]. This provided state-of-the-art results in parsing Romance languages such as Italian [35] and Spanish [26]. In the latter case, we mixed the outputs of two morphological analyzers and generated a version of the treebank where each morphological gold information was replaced by a predicted one. Relying on a rich lexicon developed within the Alexina framework (cf. 5.8) and accurate morphological treatment (cf. 6.5), this method brings more robustness to treebank-based parsing models.

6.4.3. Out-of-domain parsing : resources and parsing techniques

Statistical parsing is known to lead to parsers that exhibit quite degraded performance on input text that varies from the sentences used for training. Alpage has devoted a major effort on providing both evaluation resources and parser adaptation techniques, to increase robustness of statistical parsing for French. We have investigated several degrees of distance between the training corpus, the French Treebank, which is made of sentences from the *Le Monde* newspaper: we first focused on parsing well-edited texts, but from domains with varying difference with respect to the national newspaper *Le Monde* type of text. We then turned our attention to parsing user-generated content, hence potentially not only from a different domain than news, but also with great “noise” with respect to well-edited texts, and extremely divergent linguistic phenomena (see next subsection). As far as out-of-domain well-edited text, we have supervised the annotation and release of the **Sequoia Treebank** [47] (https://www.rocq.inria.fr/alpage-wiki/tiki-index.php?page=CorpusSequoia), a corpus of 3200 sentences annotated for part-of-speech and syntactic structure, from four subdomains : sentences from the regional newspaper *L’Est Républicain*, from the French Wikipedia, from the Europarl Corpus (European parliamentary debates), and from reports of the European Medicine Agency. We have proposed a word clustering technique, with clusters computed over a “*bridge*” corpus that couples indomain and target domain raw texts, to improve parsing performance on target domain, without degrading performance on indomain texts (contrary to usual adaptation techniques such as self-training). Preliminary experiments were performed on the biomedical domain only [67] and confirmed on the whole Sequoia Treebank [47].

6.4.4. Robust parsing of user-generated content

Until very recently out-of-domain text genres that have been prioritized have not been Web 2.0 sources, but rather biomedical texts, child language and general fiction (Brown corpus). Adaptation to user-generated content is a particularly difficult instance of the domain adaptation problem since Web 2.0 is not really a domain: it consists of utterances that are often ungrammatical. It even shares some similarities with spoken language [116]. The poor overall quality of texts found on such media lead to weak parsing and even POS-tagging results. This is because user-generated content exhibits both the same issues as other out-of-domain data, but also tremendous issues related to tokenization, typographic and spelling issues that go far beyond what statistical tools can learn from standard corpora. Even lexical specificities are often more challenging than on edited out-of-domain text, as neologisms built using productive morphological derivation, for example, are less frequent, contrarily to slang, abbreviations or technical jargon that are harder to analyze and interpret automatically.

In order to fully prepare a shift toward more robustness, we started to develop a richly annotated corpus of user-generated French text, the French Social Media Bank, which includes not only POS, constituency and functional information, but also a layer of "normalized" text[37]. This corpus is fully available and constitutes the first data set on Facebook data and the first instance of user generated content for an MRL.

Besides delivering a new data set, our main purpose here is to be able to compare two different approaches to user-generated content processing: either training statistical models on the original annotated text, and use them on raw new text; or developing normalization tools that help improving the consistency of the annotations, train statistical models on the normalized annotated text, and use them on normalized texts (before un-normalizing them).

However, this raises issues concerning the normalization step. A good sandbox for working on this challenging task is that of POS-tagging. For this purpose, we did leverage Alpage's work on MElt, a state-of-the art POS tagging system [15]. A first round of experiments on English have already led to promising results during the shared task on parsing user-generated content organized by Google in May 2012 [93], as Alpage was ranked second and third [38]. For achieving this result, we brought together a preliminary implementation of a normalization wrapper around the MElt POS tagger followed by a state-of-the art statistical parser improved by several domain adaptation techniques originally developed for parsing edited out-of-domain texts (cf. previous section).

One of our objectives is to generalize the use of the normalization wrapper approach to both POS tagging and parsing, for English and French, in order to improve the quality of the output parses. However, this raises several challenges: non-standard contractions and compounds lead to unexpected syntactic structures. A first round of experiments on the French Social Media Bank showed that parsing performance on such data are much lower than expected. This is why, we are actively working to improve on the baselines we established on that matter.

6.4.5. Precise recovery of unbounded dependencies

We focused on a linguistic phenomena known as long-distance dependencies. These are dependencies involved a fronted element that depends on a head that is potentially embedded in the clause the element is in front of. This embedding make such dependencies very hard to recover for a parser. Though this phenomena is rare, the corresponding dependencies are generally part of predicate-argument structures, and are thus very important to recover for downstream semantic applications. We have assessed the low parsing performance of long-distance dependencies (LDDs) for French, proposed an explicit annotation of such dependencies in the French Treebank and the Sequoia Treebank, and evaluated several parsing architectures with the aim of maintaining high general performance and good performance on LDDs [22]. We found that using a non-projective parser helps for LDDs but degrades overall performance, while using pseudo-projective parsing [88] which transforms in a reversible way a non-projective treebank into a projective one) is the best strategy, in order to take advantage of the better performance of projective parsers.

6.5. Computational morphology and automatic morphological analysis

Participants: Benoît Sagot [correspondant], Marion Baranes, Virginie Mouilleron, Damien Nouvel.

Since 2011 and, Alpage members have started interacting with formal morphologists for taking part in the development and implementation of new morphological models and resources. Concerning inflectional morphology, this work has led to new versions of the morphological layer of the ALEXINA formalism, to new ALEXINA lexicons for several languages of choice (Kurdish languages and German, as mentioned above, but also Maltese and Latin, see the section on ALEXINA), and to studies about the quantitative assessment of morphological complexity, currently an active area of research in morphology, have been pursued following previous work published in 2011 [109], [126]. Concerning constructional morphology (derivation, composition) and borrowings, studies and experiments have been carried out in the context of the ANR EDyLex project and that of the collaboration with *viavoo* [45], following here as well experiments carried out in 2011 [124], [115], [127].

6.6. Advances in lexical morphology and syntax

Participants: Benoît Sagot [correspondant], Laurence Danlos, Éric Villemonte de La Clergerie.

The Alexina framework (cf. 5.8) [105] has been developed and used for developing various lexicons, in particular the Le*fff*, that are used in many tools such as POS-taggers [15] and parsers.

In 2012, the new developments within Alexina have been fourfold:

- A large amount of work has been made for developing a new morphological layer to Alexina, in collaboration with a specialist of formal morphology.
- In the context of this collaboration, new Alexina lexicons have been developed with a special focus on linguistic relevance and exhaustivity within a well-defined subset of lexical entries (e.g., Latin verbs, 1st-binyan Maltese verbs).
- The development of a new large-scale NLP-oriented Alexina lexicon has been initiated, namely that of DeLex, an Alexina lexicon for German. It is currently restricted to the morphological layer (no valency information yet) but already generates 2 million inflected lexical entries. The underlying morphological grammar makes use of the new morphological layer mentioned above.
- Following previous work, merging experiments between syntactic resources and the Le*fff* [30] and comparison experiments between such resources and the Le*fff* as reference lexicon for the FRMG parser have been carried out [43]. In the latter series of experiments, the Le*fff* has proven better, or rather more suitable, that other (converted) resources.

6.7. Named Entity Recognition and Entity Linking

Participants: Rosa Stern, Benoît Sagot.

Identifying named entities is a widely studied issue in Natural Language Processing, because named entities are crucial targets in information extraction or retrieval tasks, but also for preparing further NLP tasks (e.g., parsing). Therefore a vast amount of work has been published that is dedicated to named entity *recognition*, i.e., the task of identification of named entity *mentions* (spans of text denoting a named entity), and sometimes *types*. However, real-life applications need not only identify named entity mentions, but also know which real entity they refer to; this issue is addressed in tasks such as knowledge base population with entity resolution and linking, which require an inventory of entities is required prior to those tasks in order to constitute a reference.

6.7.1. Cooperation of symbolic and statistical methods for named entity recognition and typing

Named entity recognition and typing is achieved both by symbolic and probabilistic systems. We have performed an experiment [62] for making the rule-based system NP, SxPipe's high-precision named entity recognition system developed at Alpage on AFP news corpora and which relies on the *Aleda* named entity database, interact with LIANE, a high-recall probabilistic system developed by Frédéric Béchet (LIF) and trained on oral transcriptions from the ESTER corpus. We have shown that a probabilistic system such as LIANE can be adapted to a new type of corpus in a non-supervised way thanks to large-scale corpora automatically annotated by NP. This adaptation does not require any additional manual annotation and illustrates the complementarity between numeric and symbolic techniques for tackling linguistic tasks.

6.7.2. Nomos, a statistical entity linking system

For information extraction from news wires, entities such as persons, locations or organizations are especially relevant in a knowledge acquisition context. Through a process of named entity recognition and entity linking applied jointly, we aim at the extraction and complete identification of these relevant entities, which are meant to enrich textual content in the form of *metadata*. In order to store and access extracted knowledge in a structured and coherent way, we aim at populating an ontological reference base with these metadata. We have pursued our efforts in this direction, using an approach where NLP tools have early access to Linked Data resources and thus have the ability to produce metadata integrated in the Linked Data framework. In particular, we have studied how the entity linking process in this task must deal with noisy data, as opposed to the general case where only correct entity identification is provided.

We use the symbolic named entity recognition system NP, a component of SxPipe, and use it as a mention detection module. Its output is then processed through our entity linking system, which is based on a supervised model learned from examples of linked entities. Since our named entity recognition is not deterministic, as opposed to other entity linking tasks where the gold named entity recognition results are provided, it is configured to remain ambiguous and non-deterministic, i.e., its output preserves a number of ambiguities which are usually resolved at this level. In particular, no disambiguation is made in the cases of multiple possible mentions boundaries (e.g., *{Paris}+{Hilton}* vs. *{Paris Hilton}*). In order to cope with possible false mention matches, which should be discarded as linking queries, the named entity recognition output is made more ambiguous by adding a *not-an-entity* alternative to each mention's candidate set for linking. The entity linking module's input therefore consists in multiple possible readings of sentences. For each reading, this module must perform entity linking on every possible entity mention by selecting their most probable matching entity. Competing readings are then ranked according to the score of entities (or sequence of entities) ranked first in each of them. The reading with no entity should also receive a score in order to be included in the ranking. The motivation for this joint task lies in the frequent necessity of accessing contextual and referential information in order to complete an accurate named entity recognition; thus the part where named entity recognition usually resolves a number of ambiguities is left for the entity linking module, which uses contextual and referential information about entities.

We have realized a first implementation of our system, as well as experiments and evaluation results. In particular, when using knowledge about entities to perform entity linking, we discuss the usefulness of domain specific knowledge and the problem of domain adaptation.

In 2012, improvements have been made to Nomos by combining the NP named entity detection module with LIANE, a probabilistic system developed by Frédéric Béchet (LIF) in order to better predict possible false matches. The linking step has also been enriched with the use of a more complete and autonomous knowledge base derived from Wikipedia, as well as new parameters and ranking functions for the prediction of the mention/entity alignment.

In the context of this linking task for the processing of AFP corpora and content enrichment with metadata, we conducted a deep study of Semantic Web recent developments and especially of the Linked Data initiatives in order to consider the integration of AFP metadata in these knowledge representation frameworks. On this topic as well as the enlarged view of entity linking for semantic annotation of textual content, discussions have taken place with Eric Charton (CRIM, Montréal, Canada) during 2012 Fall.

The Nomos system as well as the general process of content enrichment with metadata and reference base population has been presented at a dedicated workshop at NAACL in June 2012 (AKBC-WEKEX 2012).

6.8. Advances in lexical semantics

Participants: Benoît Sagot [correspondant], Marion Richard, Sarah Beniamine.

In 2012, several contributions to the WOLF have been finalized and/or published. In particular, various successful attempts to enhance the coverage of the WOLF have been integrated within the master resource [23], [19], [31], [24]. A more original work has also been achieved, targeted at improving the precision of the resource by automatically detecting probable outliers [32]. This latter work has been integrated within the dedicated sloWTool platform, and these outliers partly validated by Slovene students of Romance studies. In parallel, a medium-scale manual validation effort has been achieved at Alpage thanks to the work of two Master students funded by the ANR EDyLex project, which has led to the validation of a vast majority of so-called "basic" synsets, i.e., what can be expected to be the most useful part of the resource.

The result of all this work has been integrated in a preliminary first non-alpha version of the WOLF, version WOLF 1.0b.

6.9. Techniques for transferring lexical resources from one language to a closely-related one

Participants: Yves Scherrer, Benoît Sagot.

Developing lexical resources is a costly activity, which means that large resources only exist for a small number of languages. In our work, we address this issue by transferring linguistic annotations from a language with large resources to a closely related language which lacks such resources. This research activity, funded by the Labex EFL, has started in October 2012.

First results include the development of a method to create bilingual dictionaries without any parallel data, depending solely on surface form similarities and their regularities. The resulting bilingual dictionaries are used to transfer part-of-speech annotations from one language to the other. At the moment, our methods are being tested with Wikipedia texts from various languages and dialects closely related to German, such as Dutch and Pfälzisch. We plan to extend this work to data from other language groups and to other types of linguistic annotations, for instance syntactic or semantic resources.

6.10. Modelling the acquisition of linguistic categories by children

Participants: Benoît Crabbé, Luc Boruta, Isabelle Dautriche.

This task breaks in two sub-tasks: acquisition of phonemic categories, and acquisition of syntactic categories.

Although we are only able to distinguish between a finite, small number of sound categories – i.e., a given language's phonemes – no two sounds are actually identical in the messages we receive. Given the pervasiveness of sound-altering processes across languages – and the fact that every language relies on its own set of phonemes – the question of the acquisition of allophonic rules by infants has received a considerable amount of attention in recent decades. How, for example, do English-learning infants discover that the word forms [kæt] and [kat] refer to the same animal species (i.e. *cat*), whereas [kæt] and [bæt] (i.e. *cat*∼*bat*) do not? What kind of cues may they rely on to learn that [sıηkıη] and [θıηkıη] (*sinking*∼*thinking*) can not refer to the same action? The work presented in this dissertation builds upon the line of computational studies initiated by [90], wherein research efforts have been concentrated on the definition of sound-to-sound dissimilarity measures indicating which sounds are realizations of the same phoneme. We show that solving Peperkamp et al.'s task does not yield a full answer to the problem of the discovery of phonemes, as formal and empirical limitations arise from its pairwise formulation. We proceed to circumvent these limitations, reducing the task of the acquisition of phonemes to a partitioning-clustering problem and using multidimensional scaling to allow for the use of individual phones as the elementary objects. The results of various classification and clustering experiments consistently indicate that effective indicators of allophony are not necessarily effective indicators of phonemehood. Altogether, the computational results we discuss suggest that allophony and phonemehood can only be discovered from acoustic, temporal, distributional, or lexical indicators when—on average—phonemes do not have many allophones in a quantified representation of the input. This subtask has seen the Phd defense of Luc Boruta whose Phd thesis : *"Indicators of allophony and phonemehood"* was successfully defended in September 2012.

As for syntactic categorization, the task is concerned with modelling and implementing psychologically motivated models of language treatment and acquisition. Contrary to classical Natural Language Processing applications, the main aim was not to create engineering solutions to language related tasks, but rather to test and develop psycholinguistic theories. In this context, the study was concerned with the question of learning word categories, such as the categories of Noun and Verb. It is established experimentally that 2-year-old children can identify novel nouns and verbs. It has been suggested that this can be done using distributional cues as well as prosodic cues. While the plain distributional hypothesis had been tested quite extensively, the importance of prosodic cues has not been addressed in a computational simulation. We provided a formulation for modelling this hypothesis using unsupervised and semi-supervised forms of Bayesian learning (EM) both offline and online. This activity started with the master thesis of A. Gutman and has seen this year the start of a new Phd student : I. Dautriche.

6.11. Modelling and extracting discourse structures

Participants: Laurence Danlos, Charlotte Roze.

6.11.1. Lexical semantics of discourse connectives

Discourse connectives are words or phrases that indicate senses holding between two spans of text. The theoretical approaches accounting for these senses, such as text coherence, cohesion, or rhetorical structure theory, share at least one common feature: they acknowledge that many connectives can indicate different senses depending on their context. LEXCONN is a lexical database for French connectives [16].

The French connectives "*en réalité*" and "*en effet*" have been the topic of numerous studies but none of them was formalized. [53] gives a formalization of the conditions the two arguments of these connectives should meet. This formalization is based on factivity information as modeled in the FactBank corpus developed by Roser Sauri.

Sometimes, the sense of connectives is unique but its arguments are hard to determine. In particular, the second argument of an adverbial connective is not always equivalent to its syntactic arguments. This raises problems at the syntax-semantics interface which are described in [52]. The method to handle theses problems in a discursive parser will be studied in the ANR project POLYMNIE, which is headed by Sylvain Podogolla (Inria Lorraine) and which started in October 2012.

6.11.2. Discursive annotation

We plan to annotate the French corpus FTB (French Tree Bank) at the discursive level, in order to obtain the FDTB (French Discourse Tree Bank). The methodology that will be used is close to the one used in the PDTB (Penn Discourse Tree Bank). The first steps of this long term project are presented in [48], [49], [51].

This work is based on a new hierarchy of discourse relations and this new hierarchy was presented at an European workshop organized by the project MULDICO.

6.12. Modelling word order preferences in French

Participants: Juliette Thuilier, Benoît Crabbé, Margaret Grant.

We study the problem of choice in the ordering of French words using statistical models along the lines of [60] and [61]. This work aims at describing and model preferences in syntax, bringing additional elements to Bresnan's thesis, according to which the syntactic competence of human beings can be largely simulated by probabilistic models. We previously investigated the relative position of attributive adjectives with respect to the noun.

This year has seen the Phd thesis defense of Juliette Thuilier in September 2012.

In collaboration with Anne Abeillé (Laboratoire de Linguistique Formelle, Université Paris 7), we extended our corpora study with psycholinguistic questionnaires, in order to show that statistical models are reflecting some linguistic knowledge of French speakers. The preliminary results confirm that animacy is not a relevant factor in ordering French complements.

As regards to corpus work, we are extending the database with spontaneous speech corpora (CORAL-ROM and CORPAIX) and a wider variety of verbal lemmas, in order to enhance sample representativeness and statistical modelling. This activity has lead to the development of an extension of the French Treebank for oral corpora (approx 2000 sentences).

In a cross-linguistic perspective, we plan to strengthen the comparison with the constraints observed in other languages such as English or German with the recruitment of a new postdoc arriving at the beginning of 2013.

As can be seen from the outline above, this line of research brings us closer to cognitive sciences. We hope, in the very long run, that these investigations will bring new insights on the design of probabilistic parsers or generators. In NLP, the closest framework implementing construction grammars is Data Oriented Parsing (DOP).

AVIZ Project-Team

6. New Results

6.1. Tangible Visualization

Participants: Pierre Dragicevic [correspondant], Petra Isenberg, Yvonne Jansen, Jean-Daniel Fekete.

The goal of tangible visualization is to move data and controls to the physical world in order to exploit peoples' natural abilities to perceive and to manipulate objects, and to collaborate through these objects. This is a new topic in information visualization. Physical objects can be used either to represent data (physical visualizations) or to interact with data (physical controls). We studied both.

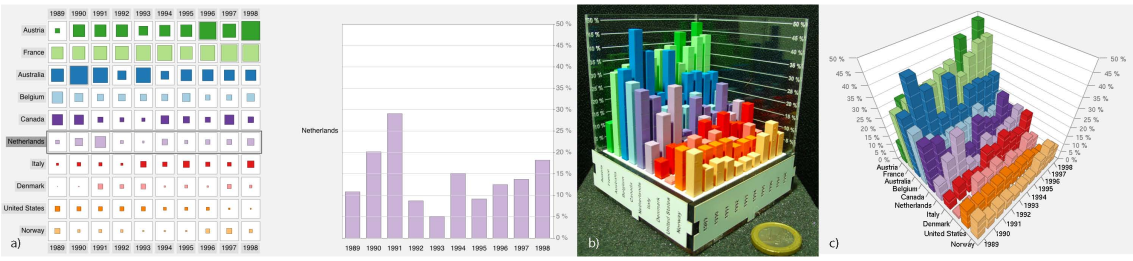


Figure 8. Education expenses of Country by Year shown under three conditions: a) on-screen 2D control; b) on-screen 3D bar chart; c) physical 3D bar chart.

Physical visualizations already exist in the form of data sculptures. Data sculptures are an increasingly popular data-driven media whose purposes are essentially artistic, communicative or educational. We are maintaining a public list of such visualizations at http://www.aviz.fr/Research/PassivePhysicalVisualizations (for passive visualizations) and at http://www.aviz.fr/Research/ActivePhysicalVisualizations (for visualizations with electronics included). But despite prolific work from the art and design communities, physical visualizations have been largely ignored in infovis research. In particular, there is no study on whether physical visualizations can help carry out actual information visualization tasks. We carried out the first infovis study comparing physical to on-screen visualizations. We focused on 3D visualizations, as these are common among physical visualizations but known to be problematic on computers. Taking 3D bar charts as an example (Figure 8), we showed that moving visualizations to the physical world can improve users' efficiency at information retrieval tasks. In contrast, augmenting on-screen visualizations with stereoscopic rendering alone or with prop-based manipulation was of limited help. Our work suggests that the efficiency of physical visualizations stem from features that are unique to physical objects, such as their ability to be touched and their perfect visual realism. These findings provide empirical motivation for current research on fast digital fabrication and self-reconfiguring materials.

We also studied how physical artifacts can help explore and interact with on-screen visualizations. One project consisted in building customizable tangible remote controllers for interacting with visualizations on wall-sized displays [34] (see http://www.aviz.fr/trc). Such controllers are especially suited to visual exploration tasks where users need to move to see details of complex visualizations. In addition, we conducted a controlled user study suggesting that tangibles make it easier for users to focus on the visual display while they interact. Another project explored the concept of stackable tangibles designed to support faceted information seeking in a variety of contexts (see http://www.aviz.fr/stackables). Each Stackable tangible represents search parameters that can be shared amongst collaborators, modified during an information seeking process, and stored and transferred. Stackables were designed to support collaborative browsing and search in large data spaces. They are useful in meetings, for sharing results from individual search activities, and for realistic datasets including multiple facets with large value ranges.

For more information, see http://www.aviz.fr/phys.

Figure 9. Four stackables. The left shows Stackables with their filter selection interface. The right two show the selected filters.

6.2. EVE : Evolutionary Visual Exploration

Participants: Evelyne Lutton [correspondant], Nadia Boukehlifa, Waldo Cancino, Anastasia Bezerianos.

Evolutionary Visual Exploration (EVE) is a new approach that combines visual analytics with stochastic optimisation to aid the exploration of multidimensional datasets characterised by a large number of possible views or projections. A prototype tool (EvoGraphDice) has been built as an extension of GraphDice, this work has been funded by the System@tics project CSDL, see Figure 10 .

Starting from dimensions whose values are automatically calculated by a PCA, an interactive evolutionary algorithm progressively builds (or evolves) non-trivial viewpoints in the form of linear and non-linear dimension combinations, to help users discover new interesting views and relationships in their data. The criteria for evolving new dimensions is not known a priori and is partially specified by the user via an interactive interface: (i) The user selects views with meaningful or interesting visual patterns and provides a satisfaction score. (ii) The system calibrates a fitness function (optimised by the evolutionary algorithm) to take into account the user input, and then calculates new views. Our method leverages automatic tools to detect interesting visual features and human interpretation to derive meaning, validate the findings and guide the exploration without having to grasp advanced statistical concepts. To validate our method, we conducted an observational study with five domain experts. Our results show that EvoGraphDice can help users quantify qualitative hypotheses and try out different scenarios to dynamically transform their data. Importantly, it allowed our experts to think laterally, better formulate their research questions and build new hypotheses for further investigation.

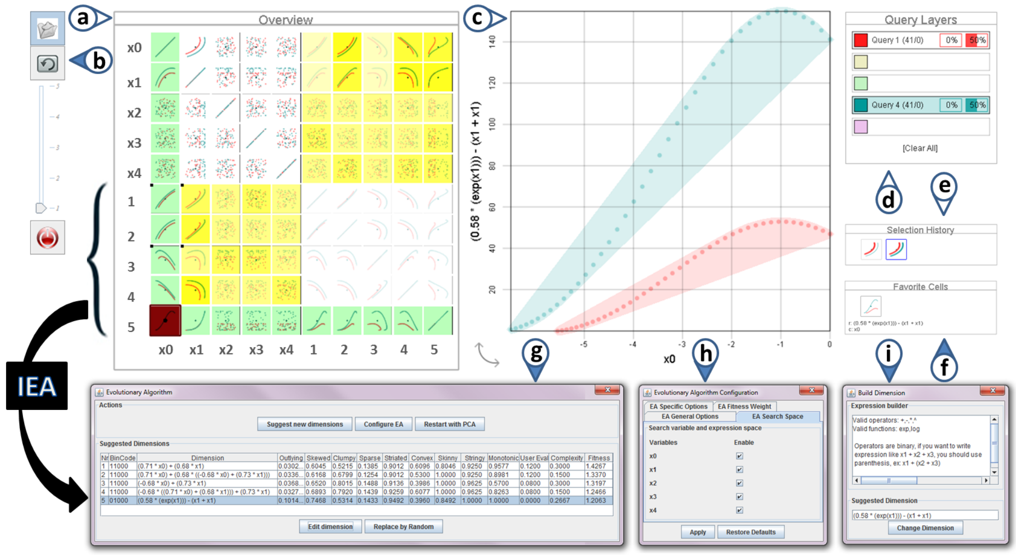


Figure 10. EvoGraphDice prototype showing an exploration session of a synthetic dataset. New extensions to the GraphDice system are indicated by coloured label arrows. Widgets: (a) an overview scatterplot matrix showing the original data set of 5 dimensions (x0..x4) and the new dimensions (1..5) as suggested by the evolutionary algorithm. (b) a tool bar with (top to bottom)"favourite" toggle button, "evolve" button , a slider to evaluate cells and a restart (PCA) button. (c) main plot view. (d) the selection query window. (e) the selection history tool. (f) the favourite cells window. (g) IEA main control window. (h) window to "limit the search space". (i) dimension editor.

6.3. Perception of Visual Variables on Wall-Sized Displays

Participants: Anastasia Bezerianos [correspondant], Petra Isenberg.

We ran two user studies on the perception of visual variables on tiled high-resolution wall-sized displays [9]. We contribute an understanding of, and indicators predicting how, large variations in viewing distances and viewing angles affect the accurate perception of angles, areas, and lengths. Our work, thus, helps visualization researchers with design considerations on how to create effective visualizations for these spaces. The first study showed that perception accuracy was impacted most when viewers were close to the wall but differently for each variable (angle, area , length). Our second study examined the effect of perception when participants could move freely compared to when they had a static viewpoint. We found that a far but static viewpoint was as accurate but less time consuming than one that included free motion. Based on our findings, we recommend encouraging viewers to stand further back from the display when conducting perception estimation tasks. If tasks need to be conducted close to the wall display, important information should be placed directly in front of the viewer or above, and viewers should be provided with an estimation of the distortion effects predicted by our work—or encouraged to physically navigate the wall in specific ways to reduce judgement error. For more information, see http://www.aviz.fr/Research/WallVariables.

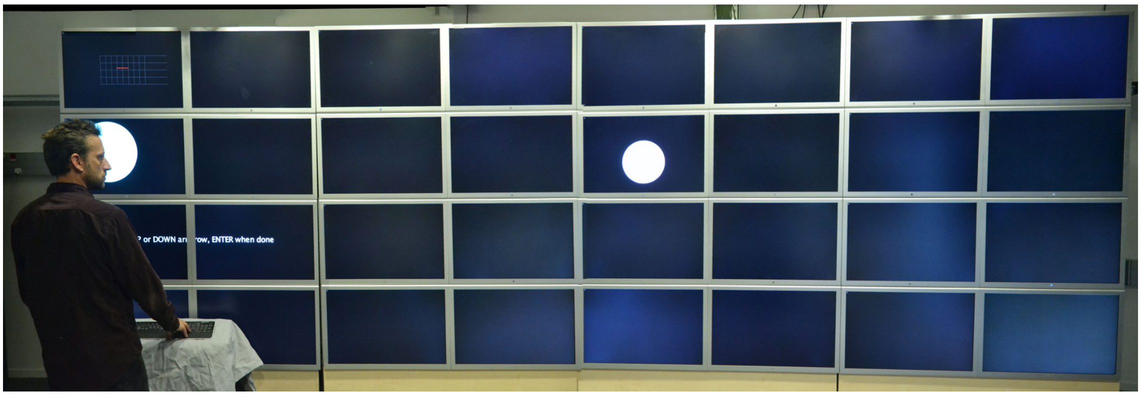


Figure 11. A participant during one trial of a user study on the WILD wall-sized display.

6.4. Sketchyness in Visualization

Participants: Tobias Isenberg [correspondant], Petra Isenberg, Jo Wood, Jason Dykes, Aidan Slingsby, Nadia Boukhelifa, Anastasia Bezerianos, Jean-Daniel Fekete.

AVIZ, in collaboration with City University London, studied how sketchiness can be used, both as a visual style and as a way to represent qualitative uncertainty.

We first studied Handy, an alternative renderer for the Processing graphics environment developed by our collaborators at the City University London [24]. It allows higher-level graphical features such as bar charts, line charts, treemaps and node-link diagrams to be drawn in a sketchy style with a specified degree of sketchiness. Our evaluation concentrated on two core aspects: the perception of sketchiness as a visual variable and higher-level impact of sketchiness on the perception of a whole graphic drawn in this style. Results suggest relative area judgment is compromised by sketchy rendering and that its influence is dependent on the shape being rendered. We showed that degree of sketchiness may be judged on an ordinal scale but that its judgement varies strongly between individuals. We evaluated higher-level impacts of sketchiness through user testing of scenarios that encourage user engagement with data visualization and willingness to critique visualization design. Results suggest that where a visualization is clearly sketchy, engagement may be increased and that attitudes to participating in visualization annotation are more positive. The results of this work have implications for effective information visualization design that go beyond the traditional role of sketching as a tool for prototyping or its use for an indication of general uncertainty.

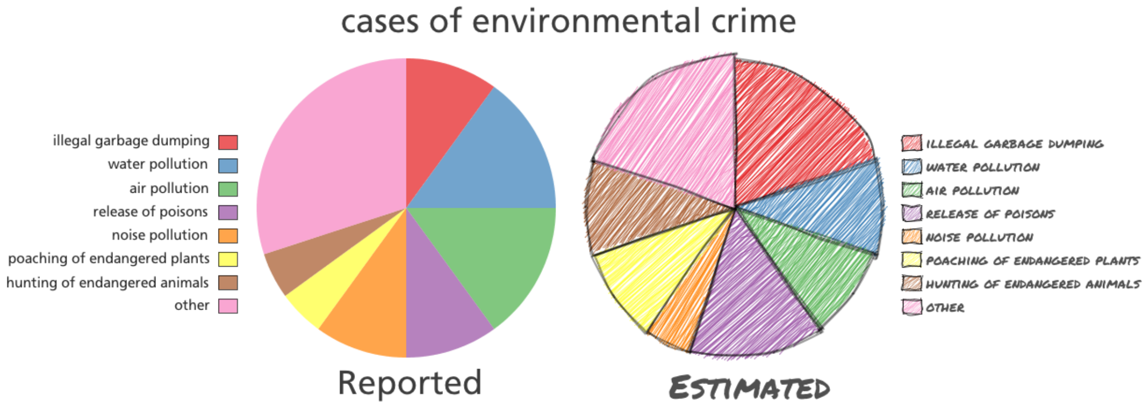


Figure 12. A pie chart drawn in regular and sketchy style.

On this last issue, we have studied whether sketchiness was an effective rendering style for conveying qualitative uncertainty [10]. We compared sketchiness to blur, intensity and dashes and obtained mixed results, showing that sketchiness is not worse than the other visual encodings but that none of them are intuitive and all of them are very limited in range, although still usable for common cases. More work is needed to asses how sketchiness can be best used and to find out more effective encodings for conveying uncertainty in a spontaneous/intuitive way.

For more information, see http://www.aviz.fr/Research/SketchyRendering and http://www.aviz.fr/Research/UncertaintySketchy.

6.5. Supporting Judgment and Decision Making with Visualizations

Participants: Pierre Dragicevic [correspondant], Luana Micallef, Jean-Daniel Fekete.

People have difficulty understanding statistical information and are unaware of their wrong judgments. Cognitive biases abound, particularly in Bayesian reasoning (see http://youtu.be/D8VZqxcu0I0 for a classic example). Psychology studies suggest that the way Bayesian problems are represented can impact comprehension, but few visual designs have been evaluated and only populations with a specific background have been involved. We conducted a study where a textual and six visual representations for three classic problems were compared using a diverse subject pool through crowdsourcing []. Visualizations included area-proportional Euler diagrams, glyph representations, and hybrid diagrams combining both. Our findings were inconsistent with previous studies in that subjects' accuracy was remarkably low and did not significantly improve when a visualization was provided with the text. A follow-up experiment confirmed that simply adding a visualization to a textual Bayesian problem is of little help for crowdsource workers. It however revealed that communicating statistical information with a diagram, giving no numbers and using text to merely set the scene significantly reduces probability estimation errors. Thus, novel representations that holistically combine text and visualizations and that promote the use of estimation rather than calculation need to be investigated. We also argued for the need to carry out more studies in settings that better capture real-life rapid decision making than laboratories. We proposed the use of crowdsourcing to partly address this concern, as crowdsourcing captures a more diverse and less intensely focused population than university students. Doing so, we hope that appropriate representations that facilitate reasoning for both laymen and professionals, independent of their background, knowledge, abilities and age will be identified. By effectively communicating statistical and probabilistic information, physicians will interpret diagnostic results more adequately, patients will take more informed decisions when choosing medical treatments, and juries will convict criminals and acquit innocent defendants more reliably.

For more information, see http://www.aviz.fr/bayes.

AXIS Project-Team

5. New Results

5.1. Introduction

As planned, our new results are splitted into our three sub-objectives as introduced below:

5.1.1. Mining for Knowledge Discovery in Information Systems

This year we get six main results: one related to how to integrate domain knowledge in a multi-view KDD process (cf. section 5.2.4), two on new KDD methods involving clustering (cf. sections 5.2.3 and 5.2.2), one on the construction of hierarchical structures of concepts in the field of e-tourism (cf. section 5.2.6), one on partitioning objects taking into account simultaneously their relational descriptions given by multiple dissimilarity matrices (cf. section 5.2.1), and finally improvment of our work on critical edition of Sanskrit texts (cf. section 5.2.5).

- Zhang based on his thesis (2010) has published this year his work on modeling and clustering users with evolving profiles in usage streams [32]. This paper will propose three models to summarize bi-streaming data, which are the batch model, the Evolving Objects (EO) model and the Dynamic Data Stream (DDS) model. Through creating, updating and deleting user profiles, the models summarize the behaviours of each user as an object. Based on these models, clustering algorithms are employed to identify the user groups. The proposed models are tested on a real-world data set showing that the DDS model can summarize the bi-streaming data efficiently and effectively, providing better basis for clustering user profiles than the other two models.
- The work described in 2011(see our AxIS annual report) on critical edition of Sanskrit texts and submitted as a paper at the Cicling 2012 conference has been accepted [21].
- A past work accepted in an international journal with A. Ciampi and colleagues [16].
- One article in an international conference on functional data analysis issued from a collaboration with F. Rossi [40].
- Two articles have been deposit in the Computing Research Repository (CoRR): one on clustering Dynamic Web Usage data [65] from A. Da Silva's thesis and one on functional data analysis [66].

5.1.2. Information and Social Networks Mining for Supporting Information Retrieval

This year, we pursued two main works on clustering methods:

- the detection of communities in a social network (graph extracted from relationnal data) (cf. section 5.3.1),
- the improvment of our dynamic hard clustering method for relational data (cf. section 5.3.2).

5.1.3. Multidisciplinary Research For Supporting User Oriented Innovation

With the expansion of the innovation community beyond the firm's boundaries (the so-called "open innovation") a lot of changes have been introduced in design and evaluation processes : the users can become co-designers, HCI design and evaluation focus is no longer placed on usability only but also on the whole user experience, experimentations take place out of lab with large number of heterogeneous people instead of carefully controlled panels of users ... All these deep changes require improvements of existing practices, methods and tools for the design / evaluation of information systems as well as for usage analysis. This evolution calls also for a structured user centered methodology (methods and ICT tools) to deal with open innovation. Various different disciplines and trends are dedicated in understanding user behavior on Internet and with Digital Technologies, notably Human Computer Interaction community (HCI), CSCW, Workplace Studies, Distributed Cognition and Data Mining. Our contribution to open innovation research keeps its focus on usage analysis for design, evaluation and maintenance of information systems and our activities this year, as indicated in our roadmap presented at the Inria theme evaluation (2011) have been conducted both breadth wise and in depth with two main objectives :

- Improving, designing and evaluation support tools for innovation,
- Development of the FocusLab platform.

The research was conducted along three focus:

- Extension of usability methods and models (cf. section 5.4),
- Designing and evaluating user experience in the context of a living lab process (cf. section 5.5),
- FocusLab Platform (cf. section 5.6).

Let us note one research work related to Living labs done in 2011 and published in 2012 [26].

5.2. Mining for Knowledge Discovery in Information Systems

5.2.1. Clustering on Multiple Dissimilarity Matrices

Participants: Yves Lechevallier, F.A.T. de Carvalho, Guillaume Pilot, Brigitte Trousse.

In [17], we introduce hard clustering algorithms that are able to partitioning objects taking into account simultaneously their relational descriptions (relations + values) given by multiple dissimilarity matrices. The aim is to obtain a collaborative role of the different dissimilarity matrices in order to obtain a final consensus partition. These matrices could have been generated using different sets of variables and a fixed dissimilarity function or using a fixed set of variables and different dissimilarity functions, or using different sets of variables and dissimilarity functions.

During 2012 we show interest and disadvantages of these approaches to classifying curves with a Urso and Vichi distance based on the mathematical properties of curves (first derivative and second). The curves are issued from temperature sensors placed in 40 offices during one year (See section 6.1.3). This period was divided into the periods before and after challenge and the challenge period. During the challenge period the occupants had information by bonus / malus messages on energy consumption [34].

5.2.2. Web Page Clustering based on a Community Detection Algorithm

Participants: Yves Lechevallier, Yacine Slimani.

Extracting knowledge from Web user's access data in Web Usage Mining (WUM) process is a challenging task that is continuing to gain importance as the size of the Web and its user-base increase. That is why meaningful methods have been proposed in the literature in order to understand the behaviour of the user in the Web and improve the access modes to information. We pursued our previous work [102] and defined a new approach of knowledge extraction using graph theory. which is described in [29].

This work is done in collaboration with the laboratory LRIA At the Ferhat Abbas University, Sétif, Algérie.

5.2.3. Multi-criteria Clustering with Weighted Tchebycheff Distances for Relational Data

Participants: F.A.T. de Carvalho, Yves Lechevallier.

The method described in [27] uses a nonlinear aggregation criterion, weighted Tchebycheff distances, more appropriate than linear combinations (such as weighted averages) for the construction of compromise solutions. We obtain a partition of the set of objects, the prototype of each cluster and a weight vector that indicates the relevance of each criterion in each cluster. Since this is a clustering algorithm for relational data, it is compatible with any distance function used to measure the dissimilarity between objects.

5.2.4. Knowledge management in Multi-View KDD Process

Participant: Brigitte Trousse.

E.L. Moukhtar Zemmouri, in the context of his PhD thesis supervised by Hicham Behja, A. Marzark and B. Trousse pursued his work based on a Viewpoint Model in the context of a KDD process [30], [19].

Knowledge Discovery in Databases (KDD) is a highly complex, iterative and interactive process aimed at the extraction of previously unknown, potentially useful, and ultimately understandable patterns from data. In practice, a KDD process (data mining project according to CRISP-DM vocabulary) involves several actors (domain experts, data analysts, KDD experts, etc.) each with a particular viewpoint. We define a multi-view analysis as a KDD process held by several experts who analyze the same data with different viewpoints.
We propose to support users of multi-view analysis through the development of a set of semantic models to manage knowledge involved during such an analysis. Our objective is to enhance both the reusability of the process and coordination between users.
To do so, we propose first a formalization of viewpoint in KDD and a Knowledge Model that is “a specification of the information and knowledge structures and functions involved during a multi-view analysis”. Our formalization, using OWL ontologies, of viewpoint notion is based on CRISP-DM standard through the identification of a set of generic criteria that characterize a viewpoint in KDD. Once instantiated, these criteria define an analyst viewpoint. This viewpoint will guide the execution of the KDD process, and then keep trace of reasoning and major decisions made by the analyst.
Then, to formalize interaction and interdependence between various analyses according to different viewpoints, we propose a set of semantic relations between viewpoints based on goal-driven analysis. We have defined equivalence, inclusion, conflict, and requirement relations. These relations allow us to enhance coordination, knowledge sharing and mutual understanding between different actors of a multi-view analysis, and reusability in terms of viewpoint of successful data mining experiences within an organization.

5.2.5. *Critical Edition of Sanskrit Texts*

Participants: Yves Lechevallier [correspondant], Marc Csernel, Ehab Assan.

With the help of Ehab Assan we improved the prototype made last year by Nicolas Bèchet (cf. 2011 AxIS activity report,[21]). It is now included in the construction process of critical editions of Sanskrit texts. Ehab also added LaTeX output to the process, we now have paper as well as Web output. It was possible to present these new features [33], [36] at the 13th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing) in Delhi.

5.2.6. *Construction and Settlement of hierarchical Structures of Concepts in E-tourism*

Participant: Yves Lechevallier.

The work of Nicolas Bechet (AxIS member in 2011) and Yves Lechevallier in collaboration with Marie-Aude Aufaure (Ecole Centrale de Paris), was published in 2012 [20] related to a method for the construction and the automatic settlement of hierarchical structures of concepts. We were particularly interested in the construction of a hierarchical structure of services offered in Hotels from a data set of an application in the field of e-tourism motivated by our contacts with the SME Addictrip. The goal is to associate to each service a concept that provides a common representation of all services. Our experiments are carried out using resources from partners specialized in online hotel booking, in particular from Addictrip. The establishment of a structure of concepts is essential to these partners that use their own terminologies description of hotel services. Indeed it provides a common representation space allowing the comparison of service coming from different resources. Our approach is based on proximity of literal terms in the service having a nearby measure based on n-grams of characters. The results during our experiments show the quality of this approach and its limitations.

5.3. Information and Social Networks Mining for Supporting Information Retrieval

5.3.1. *Clustering of Relational Data and Social Network Data*

Participants: Yves Lechevallier, Amine Louati, Bruno Almeida Pimentel.

The automatic detection of communities in a social network can provide a kind of graph aggregation. The objective of graph aggregations is to produce small and understandable summaries and it can highlight communities in the network, which greatly facilitates the interpretation.

Social networks allow having a global view of the different actors and different interactions between them, thus facilitating the analysis and information retrieval.

In the enterprise context, a considerable amount of information is stored in relational databases. Therefore, relational database can be a rich source to extract social network.

During this year many updates of the program developed by Louati Amine in 2011 were performed by Bruno Almeida Pimentel. A book chapter, included the new aggregation criteria proposed ans evaluted by Bruno Almeida Pimentel, was written and will be published in 2013.

This work is done in collaboration with Marie-Aude Aufaure, head of the Business Intelligence Team, Ecole Centrale Paris, MAS Laboratory.

5.3.2. Multi-View Clustering on Relational Data

Participants: Thierry Despeyroux, Yves Lechevallier.

In the work reported in [23] in collaboration with Francisco de A.T. de Carvalho, we introduce an improvement of a clustering algorithm described in [17] that is able to partition objects taking into account simultaneously their relational descriptions given by multiple dissimilarity matrices. In this version of the prototype clusters depend on the variables of the representation space. These matrices could have been generated using different sets of variables and dissimilarity functions. This method, which is based on the dynamic clustering algorithm for relational data, is designed to provided a partition and a vector of prototypes for each cluster as well as to learn a relevance weight for each dissimilarity matrix by optimizing an adequacy criterion that measures the fit between clusters and their representatives. These relevance weights change at each algorithm iteration and are different from one cluster to another. Moreover, various tools for the partition and cluster interpretation furnished by this new algorithm are also presented.

Two experiments demonstrate the usefulness of this clustering method and the merit of the partition and cluster interpretation tools. The first one uses a data set from UCI machine learning repository concerning handwritten numbers (digitalized pictures). The second uses a set of reports for which we have an expert classification given a priori. which we have an expert classification given a priori.

5.4. Extension of Usability Methods and Tools

5.4.1. User Evaluation and Tailoring of Personal Information

Participants: Claudia Detraux, Dominique Scapin.

In the context of the ANR project PIMI (Personal Information Management through Internet see section 6.2.1), ergonomic inspections have been carried out to evaluate the usability of the PIMI V0.1. prototype, in its PC and mobile versions [49], [48]. Also, an experiment [24], [35] was conducted on a mockup of a Personal Information Space. Users were asked to perform scenario-based data entry and retrieval tasks, then to modify the mockup according to their wishes and needs. The results allowed to validate the item content and structure for the future personal space, as well as to assess the role of user modifications as evaluation cues, and for the development of further ergonomic recommendations. Detailed information was obtained on how users enter and retrieve data, by modifying the interface settings and shows that the adaptable nature of a Personal Information Space can indeed influence its acceptance, and provides useful cues for ergonomic evaluation

5.4.2. Usability Methods for Information Visualization

Participant: Dominique Scapin.

A collaboration between UFRGS (Federal University of Rio Grande do Sul, Institute of Informatics), Brazil and Inria-AxIS led to a book chapter [37] dealing with potential methodologies for including a user-centered approach into information visualization techniques. It starts by presenting the evolution of visualization techniques evaluation, briefly summarizing the main contributions in this area since its humble beginning as a collateral activity until the recent growth of interest. Then, the focus is on current issues related to such evaluations, particularly concerning the way they are designed and conducted, taking into account a background of well-known usability evaluation methods from HCI to help understanding why there are still open problems. A set of guidelines for a (more) user-centered usability evaluation of information visualization techniques is proposed and discussed.

5.4.3. Usability Recommendations for MIS (Mixed Interactive Systems)

Participant: Dominique Scapin.

A collaboration between University of Toulouse – IUT Tarbes, IRIT and Inria-AxIS led to a book chapter [38] dealing with Mixed Interactive Systems (MIS) which denote an advanced form of interaction that aims at combining physical and digital worlds, such as mixed and augmented reality, tangible user interfaces, ubiquitous computing, etc. Their main interest relates to the use of physical artifacts from the user's activity customary context. The book chapter first reports on a systematic review of the literature on MIS evaluation. From that review, usability recommendations were selected and deciphered before reformulating them under a common format. Finally, three different classification schemes of the usability recommendations obtained are proposed to facilitate search and retrieval, but also to better integrate them into the MIS development process.

5.5. Designing and Evaluating User Experience and Methods for Open Innovation

5.5.1. From Usability to User Experience: an HCI Review

Participants: Dominique Scapin, Bernard Senach, Brigitte Trousse, Marc Pallot.

Through an extensive review of the literature, a paper [28] attempted to characterize a rather novel and popular view on human-computer interaction: User Experience (UX). After introducing its polysemous nature, this paper describes the origins of UX, its scope, underlying concepts and components, as well as its various definitions. Then, UX methods are surveyed and classified, distinguishing processes, frameworks, and specific methods. The paper identifies a set of issues about the needs for increased UX maturity. Even though UX can still be viewed as an extension of usability, its future may correspond to a paradigm evolution rather than simply a buzz word. The evolution is not drastic, but it adds complexity (including new measurements) by considering a few more user areas than traditional usability.

5.5.2. Evaluation of our Methods for Idea Generation Process

Participants: Anne-Laure Negri, Caroline Tiffon, Brigitte Trousse, Bernard Senach.

In 2011 we proposed a methodology coupling two methods [25] (GenIoT a generative method based on probes (fake sensors and/or actuators) and ALoHa! a bodystorming method for designing service concepts in the specific paradigm of the Internet of Things (IoT). In the frame of the European project ELLIOT - Experiential Living Lab for Internet Of Things -, ICT Usage Lab (cf. section 6.1.8) aims at co-creating "green" services, i.e. services based on air quality and noise measurement.

Both IoT ideation methods Aloha! and GenIoT were used for the co-creation of health related services (cf. section 6.3.1.1). The participants of the methods were Environment and Health professional. Results were very different than the workshops run with citizen in the frame of the mobility scenario in ELLIOT (see 2011 AxIS activity report). Comparison of these workshops shows that hybrid approaches –i.e. co-creation approaches mixing both real and virtual meetings are not working as well as pure face to face or pure online approaches. Moreover, GenIoT method seems to be more effective with citizen than with professional. Aloha! is effective in both groups but more efficient with professionals. However the participant experience of Aloha! is higher in the case of citizen (mainly because professionals are not used to practice creative thinking methods and do not appreciate to go out of their comfort zone).

5.5.3. Leading People Behavior Changes: Mining Evolutive Data

Participants: Brigitte Trousse, Yves Lechevallier, Guillaume Pilot, Carole Goffart, Bernard Senach.

The ECOFFICES project (cf. [62], [22] and section 6.1.3) was for AxIS project team our first step towards eco-behavior study. It provided us a very rich context to study how to analyse the evolution of the energy consuming of employees during an energy challenge. A qualitative analysis from questionnaires (before and after the challenge) has been done as well quantitative analysis. The data set for quantitative data is composed of heterogeneous data issued from around 400 sensors (temperature, presence, behavior in terms of opening doors, windows, bonus, malus, etc.). We made different studies related to data preprocessing and data analysis. In our first study [64], we cleaned the data set and selected reliable data for data analysis (only temperature of various equipments, user presence and bonus/malus points). We decided not to work with aggregated variables such as the initial ponderation (defined by partners) for the various bonus-malus rules and the energy consuming at the office level. We decided to use (office, day) as statistical unit (i.e. 9995 units) with a vectoral representation. Finally we realized that the three initial periods (before the challenge, during and after the challenge) on 379 days (2011-2012) should be in fact decomposed in five periods, due to the fact the first and the last periods were split into two subperiods (with and without heat). For the analysis, we apply for each (office,day) a first analysis on a vectorial representation of temperature with the MND method (cf. section focuslab) in order to identify the best partition of these. The MND method uses euclidean distance between each value of the vectorial representation and the prototypes are defined by the means. Second we did a clustering of these units based on bonus and malus and finally we made the correspondence between these two partitions. Three classes for (office, day) are obtained. The interpretation in terms of team relied difficult but we proposed various conclusions for a winner for managing a specific bonus, or in managing ambient temperature or in behavioral change.
In our second study [34] in collaboration with Francisco de A.T. De Carvalho, our goal was to improve the interpretation task at the office and team level by applying AxIS advanced methods. To do this, we applied our hard clustering method presented in [34] on this dataset where each office was characterized by two different representations:

- Interval representation: each office is characterized by a vector of intervals corresponding to the average, minimum and maximum of daily temperatures on the three temperature sensors during these five periods. Then the office is represented by a vector of 15 intervals and the distance used is Hausdorff distance. This classification is consistent with the partition into three classes obtained during the ECOFFICES project. The class obtained with nine eco-responsible ecoffices is the same. However, other offices are divided into two classes according to the type of heating used during the winter period. The classical method divided these offices into two clusters, one of which contains the offices using the radiators during the winter period.
- Sequential representation: Each office is characterized by a vector of 9 measures, the min, max and average of daily temperatures of the three sensors in these five periods. The values are ordinate versus time and the distance used is Urso and Vichi distance (adequate for curves). The results of this approach are quite different from the classical approach results. These results required more effort for their interpretation in collaboration by specialists.

5.5.4. Leading People Behavior Changes: a Literature Review

Participants: Bernard Senach, Anne-Laure Negri.

Our research towards eco-behavior study started with the ECOFFICES project (cf. sections 5.5.3 and 6.1.3 for more details) was recently complemented with a litterature review aiming at a deeper understanding of breaks and levers to eco behavior adoption. A first work was focused on the so-called "modal change problem", compiling methods and tools aimed at supporting people to use public transportation system rather than their personal car. A second work was initiated to get a better understanding of the role that users interface could play in encouraging people to adopt a specific behavior. This work is still in progress.
Eco mobility : prompting people to adopt public transportation mode rather than their personal car.

The first review of work conducted in the fields of Persuasive Communication, Commitment, Nudges and Persuasive Technology showed that behavioral change is a process with many steps requiring to support each step with specific means. For instance, if mass communication can support the public awareness of a problem, information is not sufficient to convince people to really change their behavior. It is necessary to push them to act and numerous well-known influence techniques are nowadays available. All recent technological development (geo localization, mobile devices, social networks) can provide very effective support for behavioral changes as far as they rely on design principles identified by research in Persuasive Technology. A presentation was done on this topic for GreenCode Forum [67] (see the video on youtube).

5.5.5. Future of Internet and User-Open Innovation for Smart Cities

Participants: Marc Pallot, Brigitte Trousse, Bernard Senach.

We pursued our work on this topic and contributed to a white paper [59] which is one of the main outcomes of the FIREBALL project [cf. section 6.3.1.2), a Coordination Action within the 7th Framework Programme for ICT, running in the period 2010-2012. The aim of this project was to bring together communities and stakeholders who are active in three areas, namely: research and experimentation on the Future Internet (FIRE); open and user-driven innovation in Living Labs; and urban development for smarter cities. The goal was to develop a common vision on how the different methodologies and concepts in these areas can be aligned for cities as playgrounds of open and user driven innovation related to the Future Internet.

The white paper addresses several aspects that are critical for understanding the ‘smart city’ concept and the current progress in this area. Based on cases studies and foresight reports we aim to shed light on how the concept of smart city is currently adopted by European Cities and what the ambitions and expectations are in using this concept. It investigates the drivers and bottlenecks that influence the transformation towards a “smart city”. Underlying approaches to smart cities are discussed, both in terms of the strategies and planning approaches. From this point of view, this paper explores the conditions that must be established to stimulate the transformation towards smart cities, and the resources that are available or should be made available such as investments in broadband networks and in smart applications, as well as in the capabilities to innovate. This also points to the changing structures and processes of innovation and city development. Interestingly, we see a tendency towards more decentralized and bottom-up approaches to planning and innovation. Innovation ecosystems are characterized by a combination of top down and bottom up initiatives, leading to networking and collaboration among stakeholders, which eventually extend to real innovation communities. Increasingly, citizens, advanced businesses and local governments act as proactive catalysers of innovation, shaping cities as “agents of change”.

5.6. FocusLab Platform

5.6.1. FocusLab platform: software part

Participants: Brigitte Trousse, Yves Lechevallier, Semi Gaieb, Xavier Augros, Guillaume Pilot, Florian Bonacina.

FocusLab v1.3 (software component) done inside the ELLIOT project (cf. section 6.3.1.1) and for the purposes of the CPER Telius (cf. section 6.1.5) corresponds to the design and the implementation of a set of web-services providing basic and advanced functionalities for data analysis and some other tools supporting the living lab process.

In this version, five data analysis web services are proposed including three generic web services: a classical linear regression and two AxIs methods:

- SMDS/SCDS [91]: SCDS (Sequence Clustering in Data Stream) is a clustering algorithm for mining sequential patterns (Java) in data streams developed by A. Marascu during her thesis. This software takes batches of data in the format "Client-Date-Item" and provides clusters of sequences and their centroids in the form of an approximate sequential pattern calculated with an alignment technique. We propose in this version to return the apparition frequency (min, max, average, slope) of a sequential pattern from data streams (SCDS algorithm) (see references

- GEAR for data streams compression [93], [91], [92], [94]: GEAR (REGLO in french) is an implementation of the history management strategy proposed in Marascu's thesis [1]. It takes a set of time series and provides a memory representation of these series based on a new principle, where salient events are important (in contrast to the recent events of decaying models) .

Other data analysis services and tools have been added for Living Labs needs. We propose also two clustering methods which must be downloaded as standalone software and used for mining data from living labs:

- ATWUEDA (Axis Tool for Web Usage Evolving Data Analysis) for Analysing Evolving Web Usage Data (Da Silva 'thesis 2009 [79], [83], [81], [82]) was developed in Java and uses the JRI library (http://www.r-project.org/). The ATWUEDA tool is able to read data from a cross table in a MySQL database, split the data according to the user specifications (in logical or temporal windows) and then apply the proposed approach in order to detect changes in dynamic environment. Such an approach characterizes the changes undergone by the usage groups (e.g. appearance, disappearance, fusion and split) at each timestamp. Graphics are generated for each analysed window, exhibiting statistics that characterizes changing points over time. This application for the next experiment of Green services use case is under study.
- MND method (Dynamic Clustering Method for Multi-Nominal Data) [90]: The proposed MND method (developed in C++ language) determines iteratively a series of partitions which improves at each step the underlying clustering criterion. The algorithm is based on: a) Prototypes for representing the classes; b) Representation space; c) Proximities (distances or similarities) between two individuals; d) Context-dependent proximity functions for assigning the individuals to the classes at each step. The clustering criterion to be optimized is based on the sum of proximities between individuals and the prototype of the assigning clusters.

  This method has been also successfully applied on Web logs in 2003. This year we improved our code and tested it on IoT data (temperature) issued from the ECOFFICES project (cf. sections 5.5.3 and 6.1.3).

The application of the services provided by FocusLab 1.3 and other AxIS data mining methods for the purposes of ELLIOT use cases and other experimental projects are under study.

AYIN Team

6. New Results

6.1. Markov random fields and Marked point processes

6.1.1. Stochastic modeling for very high resolution SAR image processing

Participants: Aurélie Voisin, Vladimir Krylov, Josiane Zerubia [contact].

This work was done in collaboration with DITEN, University of Genoa, with Dr. Gabriele Moser and Pr. Sebastiano B. Serpico with partial financial support of the French Defense Agency, DGA (http://www.defense.gouv.fr/dga/). The data were provided by the Italian Space Agency, ISA (http://www.asi.it/en) in the framework of the project "Development and validation of multitemporal image analysis methodologies for multirisk monitoring of critical structures and infrastructures (2010-2012)".

high resolution, synthetic aperture radar data, multi-sensor data, urban areas, supervised classification, hierarchical Markov random fields, statistical modeling, wavelets, textural features

The classification of remote sensing images including urban areas is relevant in the context of the management of natural disasters (earthquakes, floodings...), and allows to determine land-use and establish land cover maps, or to localise damaged areas. Given the huge amount and variety of data available nowadays, the main difficulty is to find a classifier that takes into account multi-band, multi-resolution, and possibly multi-sensor data. A minor part of our work was also dedicated to the change detection [14], still in the frame of the management of natural disasters.

We developed a supervised Bayesian classifier that combines a joint class-conditional statistical modeling and a hierarchical Markov random field. The first classification step deals with the statistical modeling for each target class (e.g. vegetation, urban, etc.) by using a finite mixture model, estimated by resorting to a modified stochastic expectation maximization (SEM) algorithm. Such a model is well-adapted to deal with heterogeneous classes, and each mixture component reflects the contribution of the different materials contained in a given class. When considering optical images, the statistics are modeled by using finite mixtures of Gaussian distributions. In the case of SAR amplitude imagery, we favor a finite mixture of generalized Gamma distributions. Then, at each considered resolution, the different input bands are statistically combined by using multivariate copulas. The second classification step relies on the integration of this statistical modeling in a hierarchical Markov random field integrated in a quad-tree structure. Such contextual classifier helps improving the robustness of the method with respect to noise, or to SAR speckle. A variety of algorithms were proposed to estimate the labels on hierarchical graphs. The consideration of a specific graph, here a quad-tree, allows to benefit from its good properties (e.g. causality) and to apply non iterative algorithms. Among the different algorithms employed in the literature, we chose to take into account an exact estimator of the marginal posterior mode (MPM). The cost function associated to this estimator offers the possibility to penalize the errors according to their number and the scale at which they occur: an error at the coarsest scale is stronger penalized than an error at the finest scale. Moreover, we introduce a prior estimation update that experimentally leads to improved results.

The first experiments were run on single-polarized, mono-resolution synthetic aperture radar (SAR) amplitude images. The challenge of the problem considered here is that the given input is at a single resolution and should be integrated in a multi-scale tree. Thus, we extract an extra information in the form of a multi-scale wavelet decomposition from the initial image. Then, at each level, a textural feature map (e.g. Haralick's variance) is obtained from each image in the decomposition stack, and integrated as an additional information that aims at discriminating the urban areas. Finally, at each level, the wavelet image is combined with the textural image by using copulas, as described previously. These results were presented in [10], [23].

Such a classifier is sufficiently flexible to take into account different types of data [21], [22]. Thus, we also tested coregistered data of a given area acquired at different resolutions (e.g., multiresolution SAR images), directly integrated at the different levels of the hierarchical tree. The classification of multisensor (optical/SAR) data is illustrated in Fig. 1 . In this specific example, we consider a GeoEye acquisition (resolution: 65 centimeters) and a coregistered COSMO-SkyMed SAR acquisition (resolution: 2.5 meters) of the Port-au-Prince quay (Haiti). Spatially disjoint training and test areas were manually annotated. The classification is done following 5 classes: urban areas, natural landscape, sand, containers and wet areas.

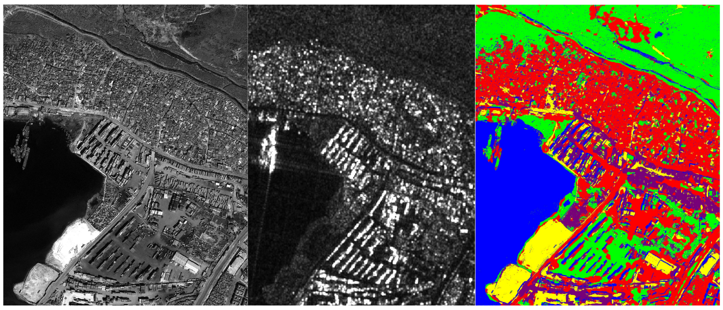


Figure 1. Left: Initial optical image of Port-au-Prince (Haiti) (©GeoEye, 2010). Middle: Initial SAR image of Port-au-Prince (Haiti) (©ISA, 2010). Right: Classification map obtained with the proposed hierarchical method for the 5 classes (Blue: wet areas; Green: vegetation; Red: urban areas; Yellow: sand; Purple: containers).

We have also run experiments on other types of acquisitions, such as histological images [22], to prove the robustness of the proposed algorithm with respect to different image sources.

6.1.2. Satellite image classification using Bootstrap EM

Participants: Siddharth Buddhiraju, Josiane Zerubia [contact].

This program has been partially funded by the Direction of International Relations of Inria (DRI).

Bootstrapping, Expectation-Maximization Algorithm, Iterated Conditional Expectation, Markov Random Fields, Simulated Annealing.

We implemented both Bootstrap EM and Iterated Conditional Expectation algorithms for parameter estimation of first order Markov Random Field models followed by Simulated Annealing, for optimal segmentation of gray-scale images. The objective was to perform a quantitative comparison of the two methods. Apart from successful implementation of these algorithms, an extension of these to multispectral images was performed, and the obtained results were found to be of superior quality compared against the original gray-scale ones (see Fig. 2).

First, Bootstrapped EM or Iterated Conditional Expectation were performed based on the user's requirement. The estimated parameters were then used to obtain the optimal segmentation of the image via simulated annealing. The algorithm was extended using multivariate Gaussian models to perform the same for multispectral images.

6.1.3. Boat detection and counting in Mediterranean Harbors using Marked Point Processes

Participants: Paula Craciun, Josiane Zerubia [contact].

This work has been conducted in collaboration with ACRI-ST (http://www.acri-st.fr/) and the French Space Agency (http://www.cnes.fr/), which provided the high resolution satellite images.

stochastic modeling, marked point process, object extraction, SEM, simulated annealing

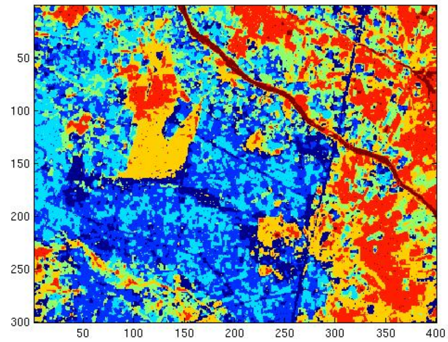


Figure 2. A sample result for a 4-band IRS-1A satellite image of a village in India. 7 clusters were considered to segment the image using Bootstrapped EM. The image was provided by Prof. Krishna Mohan Buddhiraju of CSRE, Indian Institute of Technology Bombay.

Marked point processes have been successfully applied in image processing analysis, when dealing with high resolution images in the purpose of feature extraction. The objective of this research was to improve the already existing marked point process model of ellipses to better fit the detection of boats in a harbor. The model involved two types of energy terms: a data term, used to determine the fidelity to the existing data (i.e. image) and a prior energy term, used to describe relationships between the objects. We proposed new energy components to model boat detection. The proposed model relied on a high number of parameters. While most of these parameters had an intuitive meaning and could be, thus, set manually, others were difficult to determine. We therefore used a parameter estimation method, based on the Stochastic Expectation - Maximization (SEM) algorithm, which proved to provide good results when combined with marked point processes. Furthermore, we proposed additional automatic procedures based on mathematical morphology to determine critical parameters of this model. Experimental results of boat detection are shown on Fig. 3 .

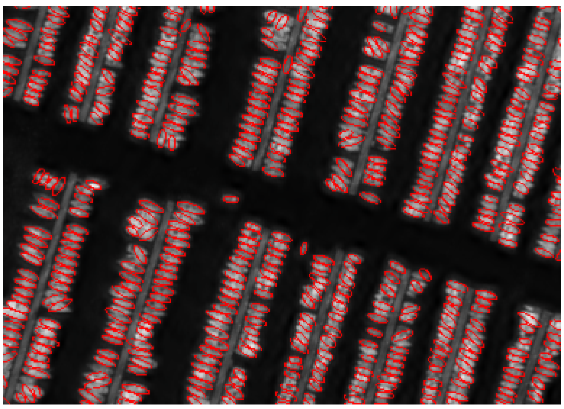


Figure 3. Result of boat detection using Marked Point Processes.

6.1.4. Contribution of object recognition on forest canopy images to the building of an allometric theory for trees and natural, heterogeneous forests

Participants: Jia Zhou, Josiane Zerubia [contact].

This work was done in collaboration with Xavier Descombes (Morpheme team, Inria-SAM), Dr. Pierre Coureron and Dr. Christophe Proisy at IRD, UMR AMAP (http://amap.cirad.fr/), Montpellier.

object detection, marked point processes, tree crowns, forest structure

This work aimed at providing information on the forest structure through the analysis of canopy properties as described by the spatial distribution and the crown size of dominant trees. Our approach was based on the Marked Point Processes (MPP) theory, which allows modeling tree crowns observed in remote sensing images by discs belonging to a two dimensional space. The potential of MPP to detect the trees crowns automatically was evaluated by using very high spatial resolution optical satellite images of both *Eucalyptus* plantations in Brazil and mangrove forest in French Guyana. LIDAR and simulated reflectance images were also analyzed for the mangrove application. Different adaptations (parameter settings, energy models) of the MPP method were tested and compared through the development of quantitative indices that allowed comparison between detection results and tree references derived from the field, photo-interpretation or the forest mockups.

In the case of mangroves, the estimated crown sizes from detections were consistent with the outputs from the available allometric models (Fig. 4 (Left and Middle)). Other results indicated that tree detection by MPP allowed mapping the local density of trees of young *Eucalyptus* plantations (Fig. 4 (Right), [11]) even if crown size was close to the image spatial resolution (0.5 m). However, the quality of detection by MPP decreased with canopy closeness. To improve the results, further work may involve MPP detection using objects with finer shapes and forest data measurements collected at the tree plant scale.

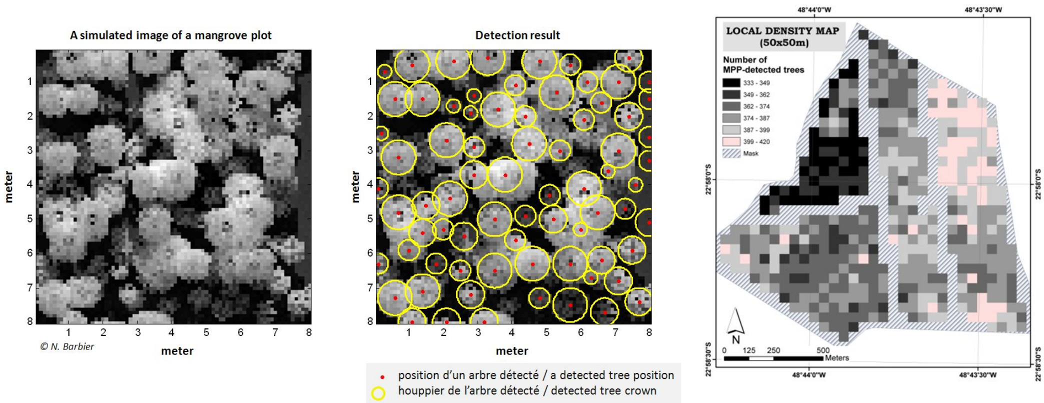


Figure 4. (Left and Middle) Result of detection on an image of mangrove. (Right) Example of a local tree density map computed over the entire plantation at 50 m resolution of the marked point process modeling. Masked areas are centered along the road network.

6.1.5. Detection of the hyperpigmentation of the skin on color images using Marked Point Process and Mathematical Morphology

Participants: Adrien Lacage, Josiane Zerubia [contact].

The source images were provided by the AYIN team itself for the study of folliculitis, and by an industrial leader in skin care for acne.

skin care, hyperpigmentation, acne, folliculitis, marked point process, mathematical morphology

Automatic detection of the skin hyperpigmentation helps in estimating the severity of some skin diseases like *acne vulgaris* and *folliculitis*. We compared two methods for studying acne and folliculitis lesions and hyperpigmentation of the skin. We adapted a model based on Marked Point Processes and initially developed for flamingo's population counting to dermatological images of acne and folliculitis. Then, we developed an algorithm which uses mathematical morphology together with volume and shadows compensation. Finally, we compared results in term of detection accuracy.

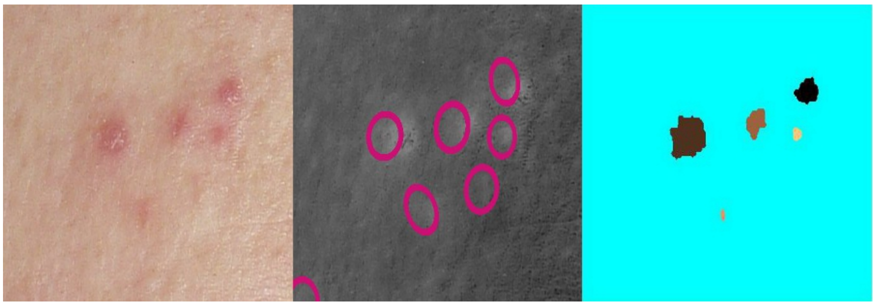


Figure 5. Source image of folliculitis (Left) and results obtained with Marked Points Processes (Middle), then with Mathematical Morphology (Right)

6.1.6. Efficient Monte Carlo sampler for detecting parametric objects in large scenes

Participants: Yannick Verdie, Florent Lafarge [contact].

This work is supervised by Florent Lafarge (Geometrica team, Inria-SAM) in collaboration with the AYIN team.

Point processes have demonstrated both efficiency and relevance when addressing object recognition problems in vision. However, simulating these mathematical models is a difficult task, especially on large scenes. Existing samplers suffer from average performances in terms of computation time and stability. We propose a new sampling procedure based on a Monte Carlo formalism. Our algorithm exploits Markovian properties of point processes to perform the sampling in parallel. This procedure is embedded into a data-driven mechanism such that the points are non-uniformly distributed in the scene. The performances of the sampler are analyzed through a set of experiments on various object recognition problems from large scenes, and through comparisons to the existing algorithms.

6.2. Statistical methods

6.2.1. Change detection on synthetic aperture radar images based on hypothesis testing

Participants: Vladimir Krylov, Josiane Zerubia [contact].

This work was conducted in collaboration with DITEN, University of Genoa with Dr. Gabriele Moser and Prof. Sebastiano Serpico (http://spt.dibe.unige.it/) with the support of the Italian Space Agency, ASI (http://www.asi.it/en).

Change detection, synthetic aperture radar, hypothesis test, likelihood ratio test, high resolution

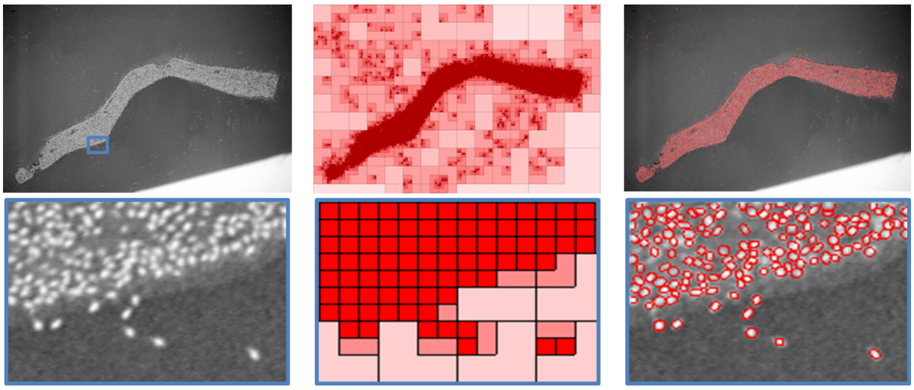


Figure 6. Bird counting by a point process of ellipses. (right) More than ten thousand birds are extracted by our algorithm in a few minutes from (left) a large scale aerial image. (middle) A quad-tree structure is used to create a non-uniform point distribution. Note, on the cropped parts, how the birds are accurately captured by ellipses in spite of the low quality of the image and the partial overlapping of birds.

Modern synthetic aperture radar (SAR) sensors represent an essential source of all-weather and 24-hour imagery with a fixed re-visit cycle at competitive high resolution. Two-date change detection from SAR images is a process that employs two SAR images acquired over the same geographical area with possibly the same (or close) acquisition characteristics at two different times to map the areas where changes occur between the two acquisition dates. The central disadvantage of the SAR imagery is given by an inherent multiplicative speckle noise, which restricts the direct application of optical-based change detection methods to SAR imagery.

We have developed a non-parametric statistical change detection approach. We avoided the ambiguity of choosing a restrictive clutter model by assuming no specific probability distribution function model [25] for the statistics of SAR. We developed a modified hypothesis test which is based on the classical Wilcoxon two-sample test that verifies whether one of two samples of independent observations tends to have larger values than the other. The choice of the Wilcoxon statistic as compared to the other available goodness-of-fit test statistics, such as, e.g., that of Cramér-von Mises' test, is a compromise solution to have simultaneously an analytically tractable asymptotic distribution (which is needed to formulate the likelihood ratio test) and a non-parametric testing procedure. Furthermore, the experimental validation demonstrated the adequacy of this statistic to the considered problem. To be able to take a decision at each pixel of the coregistered image pair we considered samples originating from the local windows centered in each pixel. Finally, we constructed a likelihood ratio test on the image with Wilcoxon statistic values. This formulation allowed to overcome the limitation of a classical independency assumption for the Wilcoxon test which is violated (at least, locally) with the local window samples. The resulting technique is related to the statistical false discovery rate approach developed for "large-scale simultaneous hypothesis testing" problems, however the derivation and interpretation are different.

Encouraging detection results were obtained on XSAR and very high resolution COSMO-SkyMed images [14].

6.2.2. Statistical analysis of skin pigmentation under treatment

Participants: Sylvain Prigent, Xavier Descombes, Josiane Zerubia [contact].

This work was partially funded by a contract with Galderma R&D (http://www.galderma.com/RampD.aspx). It was a collaboration between AYIN (Josiane Zerubia) and Morpheme (Xavier Descombes) teams.

multispectral imaging, skin, hyperpigmentation, hypothesis tests, statistical inferences

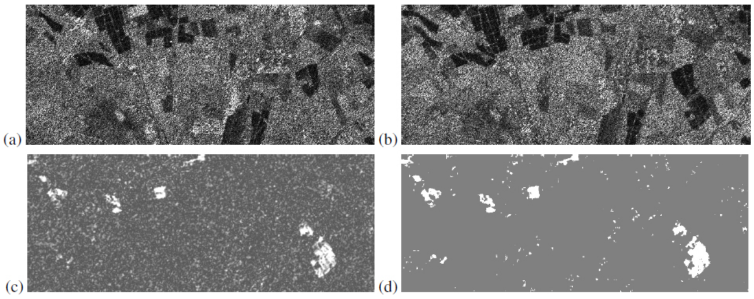


Figure 7. Coregistered XSAR images (©Univ. of Pavia) acquired on (a) April 16, 1994 and (b) April 18, 1994 and the detection results: (c) 5×5 *window-based image ratio, (d) proposed method with* 5×5 *window.*

One of the steps to evaluate the efficacy of a therapeutic solution is to test it on a clinical trial involving several populations of patients. Each population receives a studied treatment and a reference treatment for the disease. For facial hyper-pigmentation, a group of N_e patients receives the treatment on one cheek and a comparator on the other. The comparator can be a reference treatment or a placebo. To this end patients are selected to have the same hyper-pigmentation severity on the two cheeks. Then multi-spectral images are taken at different time t along the treatment period.

We proposed a methodology to estimate the efficacy a treatment by calculating three differential criteria: the darkness, the area and the homogeneity. The darkness measures the average intensity of the disease on a gray scaled image I obtained by a linear combination of the spectral bands of the original multi-spectral image. A differential darkness is then obtained by measuring the deviation between the initial measurement at time t_0, and the measurement at time t_k. The differential area criterion is calculated by analyzing the histogram of $I_{diff} = I_{t_0} - I_{t_k}$ a difference gray scale image between two measurements in a time series. The differential homogeneity criterion is obtained with a multi-scale analysis of I_{diff} adapted from the Statistical Parametric Mapping (SPM) methodology. Indeed, statistical inferences allow to assign a probability of change to each region of I_{diff} above a set of thresholds. These probabilities are calculated with respect to the maximum intensity and the spatial extend of each region. An integration of the obtained statistical map denoted SM, allows to get a homogeneity criterion.

The Fig. 8 illustrates the differential score calculated on a patient whose pathology decreases during the clinical trial. The proposed differential score has been tested in a full clinical study and provided results that agreed with the clinical analysis. This work have been patented, submitted to ISBI'13 conference and to the IEEE TMI journal, and published in Inria research reports [26], [27].

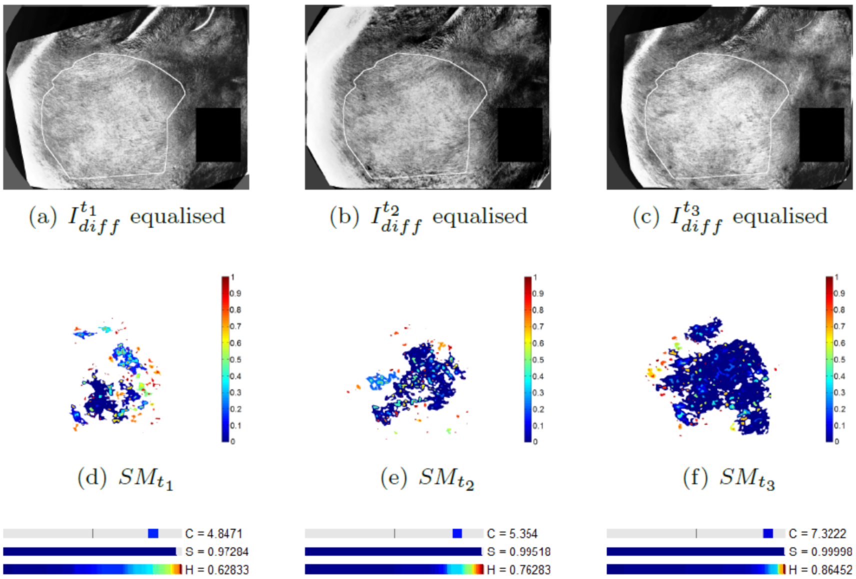
6.3. Hierarchical models

6.3.1. Hierarchical and graph cut-based models for multiyear sea ice floe segmentation

Participant: Yuliya Tarabalka [contact].

This work has been done in collaboration with Dr. Guillaume Charpiat (STARS team, Inria-SAM), Dr. Ludovic Brucker (NASA GSFC, USA) and Dr. James Tilton (NASA GSFC, USA).

hierarchical model, graph cut, segmentation, multiyear sea ice floes, shape analysis



(a) $I_{diff}^{t_1}$ equalised (b) $I_{diff}^{t_2}$ equalised (c) $I_{diff}^{t_3}$ equalised

(d) SM_{t_1} (e) SM_{t_2} (f) SM_{t_3}

(g) score t_1 (h) score t_2 (i) score t_3

Figure 8. I_{diff}, SM and differential score for the three measurements t_1, t_2, t_3 calculated for a patient whose disease decrease.

The melting of sea ice is correlated to increases in sea surface temperature and associated climatic changes. Therefore, it is important to investigate how rapidly sea ice floes melt. We proposed two methods for segmentation of a time series of a melting sea ice floe. The first method employs hierarchical model for ice floe segmentation. Image features are extracted using morphological operators, and the floe of interest is marked based on AMSR-E satellite measurements. Then, hierarchical step-wise optimization segmentation is performed, by iteratively merging adjacent regions with the smallest dissimilarity criterion. We proposed to use area and shape parameters of the floe at two previous time moments as priors for computing a segmentation map at the next time moment.

Fig. 9 (a) depicts a graph of the multiyear ice floe area as a function of time, computed by applying the proposed hierarchical model to the summer series of Moderate-Resolution Imaging Spectroradiometer (MODIS) images. While a multiyear ice floe can only melt in the summer period, peaks on the graph correspond to segmentation errors, which are a consequence of either a cloud cover or weakness of contrast between the multiyear ice and the surrounding young ice floes. These segmentation imprecisions can be avoided by simultaneously optimizing all segmentation maps in a time series. For this purpose, we developed a new method based on graph cuts for joint segmentation of monotonously shrinking (or growing) shapes. We impose shape shrinkage (or growth, respectively) constraint in graph cuts, and minimization of energy computed on the resulting graph of the image sequence yields globally optimal segmentation. Fig. 9 (c-d) show examples of floe segmentations using the new approach. Fig. 9 (b) presents a graph of the floe area as a function of time computed by performing the proposed graph cut-based method. The results are compared to those obtained by applying graph cut segmentation to each single image in the considered time series. It can be seen that the new approach yields results with continuous shrinkage of the shape size.

6.3.2. Hierarchical model for spectral-spatial classification of hyperspectral images

Participant: Yuliya Tarabalka [contact].

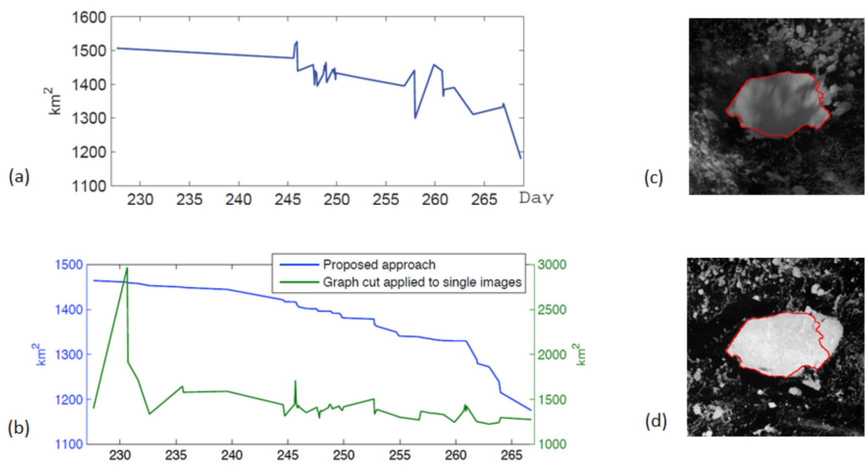


Figure 9. (a) Area of a multiyear ice floe as a function of time, computed by applying the proposed hierarchical model. (b) Area of the floe as a function of time (days), computed by using the proposed graph cut-based model (blue) and by applying graph cut segmentation to each single image in a time series. (c-d) Examples of floe segmentations using the proposed graph cut-based model.

This work has been done in collaboration with Dr. James Tilton (NASA GSFC, USA).

hyperspectral images, classification, segmentation, geometrical features, rectangularity.

The recent advances in hyperspectral remote sensor technology makes it possible to acquire data with a very high spectral (hundreds of spectral channels) and spatial (order of a meter) resolution. The rich spectral information of the hyperspectral data leads to the potential of a more accurate classification, but also presents challenges of high-dimensional data analysis.

We developed a new method for spectral-spatial classification of hyperspectral images. The method is based on the integration of probabilistic classification and shape analysis within the hierarchical step-wise optimization algorithm. First, probabilistic support vector machines classification is applied. Then, at each iteration two neighboring regions with the smallest dissimilarity criterion are merged, and classification probabilities are recomputed. We proposed to estimate a dissimilarity criterion between regions as a function of statistical, classification and geometrical (area and rectangularity) features. Fig. 10 shows the obtained classification results on a 102-band ROSIS image of the Center of Pavia, Italy, which are compared with Support Vector Machines (SVM) classification results. These results did show that the proposed method succeeded in taking advantage of both spatial and spectral information for accurate hyperspectral image classification.

6.3.3. Classification of combined hyperspectral and panchromatic data using spectral-spatial approaches

Participants: Yuliya Tarabalka [contact], Josiane Zerubia.

This work has been conducted in collaboration with the French Space Agency CNES (http://www.cnes.fr/), with Dr. Marie-José Lefèvre, Dr. Hélène DeBoissezon and Mr. Manuel Grizonnet.

hyperspectral data, HYPXIM, data fusion, panchromatic image, segmentation

Hyperspectral imaging records a detailed spectrum for each pixel, opening new perspectives in classification. Currently, several hyperspectral satellite missions such as EnMAP (210 bands, GSR 30m) are under development. The future hyperspectral satellite missions PRISMA and HYPXIM also include a panchromatic channel

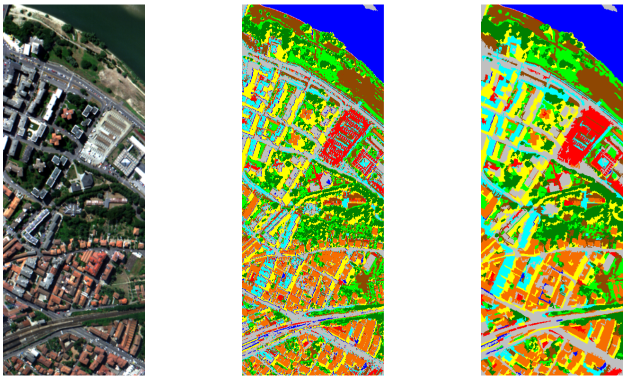


Figure 10. Center of Pavia image. (Left) Three-band color composite. (Middle) SVM classification map, overall classification accuracy = 94.9%. (Right) Classification map obtained using the proposed hierarchical approach, overall classification accuracy = 97.1%.

with better spatial resolution. We explored if a panchromatic channel at a higher spatial resolution (factor 4) contributes for more accurate classification of hyperspectral images in space conditions.

We adapted and compared several classification methods for combined hyperspectral and panchromatic images, and conducted experiments on the simulated HYPXIM data provided by CNES. We fused both data sources using principal component and Gram-Schmidt fusion methods, as well as the vector stacking approach. We then applied Support Vector Machines (SVM) classification on the resulting feature sets. Furthermore, we considered spatial information for more accurate classification by: (1) including Haralick's texture features in the feature set; (2) segmenting an image into homogeneous regions using a Hierarchical Step-Wise Optimization (HSWO) technique, and assigning each segmented region to the dominant class within this region.

Classification results are illustrated in Fig. 11 . We concluded that classification accuracies of the HYPXIM simulated data have been improved when including a panchromatic channel at a higher spatial resolution into a classification system. These results are close to hyperspectral aerial data classifications. For the image containing one-pixel regions and mixed pixels, standard spectral-spatial classification methods are not well adapted and thus do not improve accuracies when compared to pixelwise classification. In the future, we plan to develop methods which would use both spatial information and a spectral unmixing concept for efficient fusion of hyperspectral and panchromatic data.

6.4. Other detection approaches

6.4.1. Multiple-instance object detection using a higher-order active shape prior

Participants: Ikhlef Bechar, Josiane Zerubia [contact].

This work is done in collaboration with Dr. Ian Jermyn of Durham University (United Kingdom, https://www.dur.ac.uk/mathematical.sciences/) and was funded by a contract with the EADS foundation (http://www.fondation.eads.com/).

object detection, shape prior, transformation invariance, higher-order active contours, energy minimization, non-convex energy, exact convex relaxation.

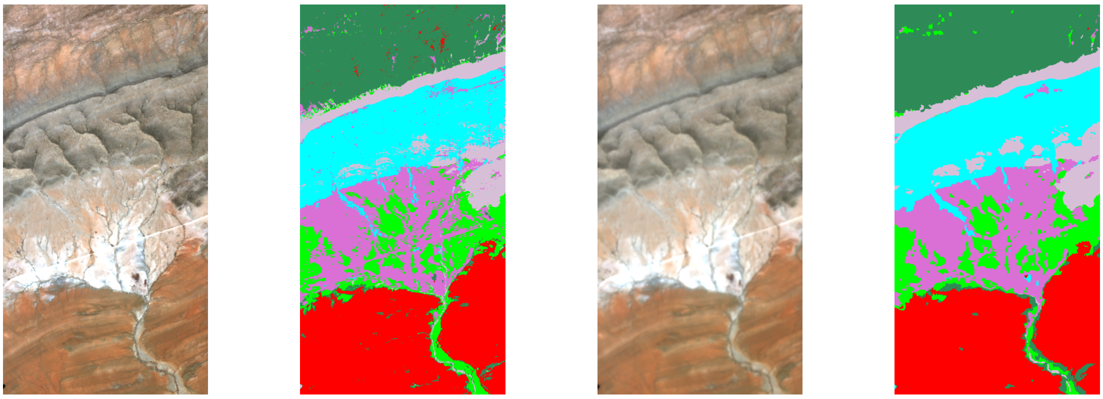


Figure 11. From left to right (a-d): (a) The false-color HYMAP aerial image (126 bands, GSR 4.8m). (b) SVM classification map for the HYMAP image, overall classification accuracy $= 83\%$. *(c) Simulated HYPXIM image (126 bands, GSR 14.4m). (d) Classification map of the fused by vector stacking hyperspectral image (c) and panchromatic image at GSR 4.8m, using HSWO-based spatial regularization, overall classification accuracy* $= 80.7\%$.

The problem under consideration is the multiple-instance object detection from imagery using prior shape knowledge. As mathematical and algorithmic framework, we have used the higher-order active contour (HOAC) model framework in order to incoporate prior shape knowledge about a class of objects of interest. On top of its robustness and its computational attractiveness (due to its parameter-estimation free method), the HOAC object-detection framework allows to incorporate shape knowledge about multiple occurrences of an object of interest in an image and to carry out object detection in a single algorithmic framework via the minimization of energy of the form:

$$\min_{\text{over all shapes}\gamma} E(\gamma) = E_{image}(\gamma) + E_{prior}(\gamma) \tag{32}$$

where γ stands for the contour an image object, $E_{image}(\gamma)$ stands for its image-based energy and $E_{prior}(\gamma)$ stands for a prior energy which is only a function of an objet's shape (and not of image data). The goal of this project is thus to model $E_{prior}(\gamma)$ using the HOAC methodology.

In this work, we have developed a fourth-order active contour (FOAC) framework for incorporating prior shape knowledge about target shapes. Typically, we express a FOAC energy model as

$$E_{foac}(\gamma) = \lambda_C L(\gamma) + \alpha_C A(\gamma) + \beta_C \int\int\int\int \langle \dot{\gamma}_p, \dot{\gamma}_q \rangle \, \mathbf{K}_C \left(|\gamma_p - \gamma_q|, |\gamma_s - \gamma_t| \right) \langle \dot{\gamma}_s, \dot{\gamma}_t \rangle \; dp\, dq\, ds\, dt \tag{33}$$

where $L(\gamma)$ and $A(\gamma)$ stand respectively for the length and the area of a contour γ and $\int\int\int\int \langle \dot{\gamma}_p, \dot{\gamma}_q \rangle \, \mathbf{K}_C \left(|\gamma_p - \gamma_q|, |\gamma_s - \gamma_t| \right) \langle \dot{\gamma}_s, \dot{\gamma}_t \rangle \; dp\, dq\, ds\, dt$ models fourth-order interactions between quadruples of contour points, and λ_C, α_C and β_C stand for some tradeoff parameters that control the contribution of each term of the FOAC energy. Note that the parameters of the method include both the real coefficients λ_C, α_C and β_C and the bivariate kernel $\mathbf{K}_C(u, v)$; $\forall\, u, v \in \mathbf{R}^+$. These parameters need to be tuned optimally for a given target shape γ^*. Thus we have developed a direct method for the optimal estimation of the FOAC parameters.

We have then shown that shapes with arbitrary geometric complexity can be modeled using such the FOAC framework 2 , and we have developed a direct method for the estimation of the parameters for a given class of shapes. In order to be able to detect multiple occurrences of a target object in an image, one needs to re-express such an originally contour-based energy 2 by replacing appropriately in formula 2 the one-dimensional contour quantity γ with an equivalent two-dimensional quantity (ie. with respect to the image domain) such as the characteristic function of γ and to minimize with respect to it the resulting energy functional. This allows topological changes of an evolving contour and hence the detection of possible multiple instances of a target object in an image. We have shown that such a new formalism is a third-order Markov Random Field (MRF) which practical optimization was a challenging question. Therefore, we have also developed a method for the exact minimization of the energy of the resulting MRF model (using a equivalent convex-relaxation approach, see Fig. 12).

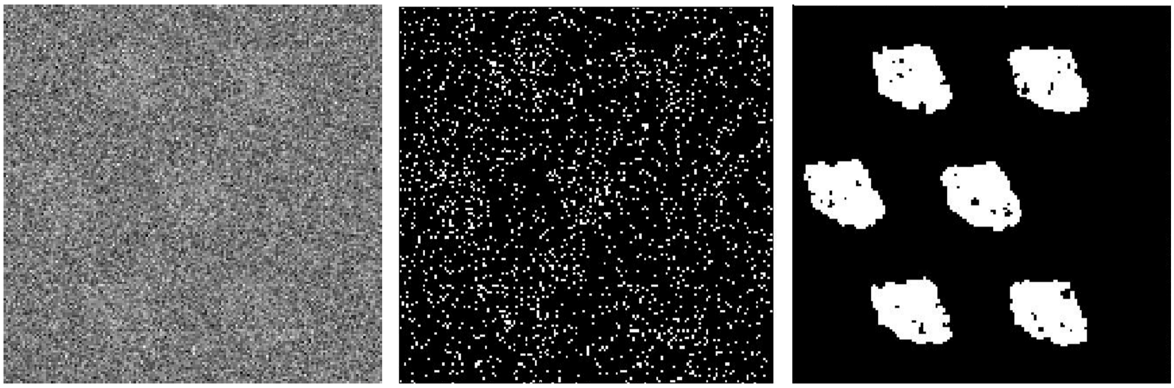


Figure 12. (Left) A very noisy input image; Multiple-instance object detection using: (Middle) a traditional segmentation method (Mumford-Shah model with a TV-based regularization) alone; (Right) with a FOAC shape prior.

6.4.2. Image analysis for automatic facial acne detection and evaluation

Participants: Zhao Liu, Josiane Zerubia [contact].

This work is part of LIRA Skin Care Project, which includes four key partners: Philips R&D (Netherlands, http://www.philips.nl), CWI (Netherlands, http://www.cwi.nl/), Fraunhofer Institutes (Germany, http://www.fraunhofer.de/en.html) and Inria (France).

image processing, feature extraction, pigmentation distributions, acne, cosmetology

Acne vulgaris is a highly prevalent skin disease, which has a significant life quality impact on sufferers. Although acne severity is readily observed by the human eye, it is an extremely challenging task to relate this visual inspection to measurable quantities of various skin tones and types. So far there is no golden standard for acne diagnosis in clinics, and it entirely depends on dermatologists' experience for evaluation of acne severity. But significant inter-rater variability among individual assessment may lead to less trustworthy diagnosis when several clinicians get involved in the study. In addition, less reproducibility of human evaluation makes comparison of acne changes over time difficult. Therefore, the long-term objective of this study is to construct an automatic acne grading system through applying spectroscopy imaging techniques and image processing methods, to objectively evaluate severity of skin disorder. Such a computer-based tool would also significantly benefit the development of better skin care products, if it can reliably characterize treatment effects of products in individual skin layers in agreement with physiological understanding.

Acne segmentation is normally considered as the first significant step in an automatic acne grading system, because segmentation accuracy directly influences the definition of acne pigmentation level, what has an impact on the goodness of acne severity evaluation. An initial unsupervised segmentation method is proposed for conventional RGB images, whose process is demonstrated in Fig. 13 (a). After several pre-processing steps (background and skin hair removal, illumination corrections), nine pigmentation descriptors were extracted from three RGB channels based on colorimetric transformations and absorption spectroscopy of major chromophores. It has been proved that the derived hemoglobin, normalized red, and normalized green descriptors can properly characterize pigmentation distributions of acne, and they are used as segmentation features. Finally, an iterative unsupervised segmentation was performed to maximize pigmentation distributions between acne and normal skin. Fig. 13 (b) shows an example of acne image on human face captured by a conventional RGB camera, while experimental result in Fig. 13 (f) illustrates that suspicious acne areas and healthy human skin can be automatically discriminated by applying the proposed method. Moreover, it only takes 90.8 seconds to segment the example image with the size of 640×428 pixels, which demonstrates the computation efficiency of the algorithm.

It should be noted that the segmentation method stated above is an initial approach. Shadows around non-flatten areas on human face (e.g. areas around nose) have a large influence on accuracy of automatic acne detection. However, based on the initial experimental results, it is difficult to entirely get rid of these effects using RGB channels only. Our finding is actually consistent with the existing studies, where researchers divided human face into several sub-regions and worked on these sub-regions individually to avoid shadow influence. Therefore, the next step study will compare acne segmentation results derived from RGB images and multi- or hyperspectral images, to investigate the most effective bands for describing acne pigmentation, as well as whether the introduction of multi- or hyperspectral analysis to the automatic acne detection and evaluation is necessary.

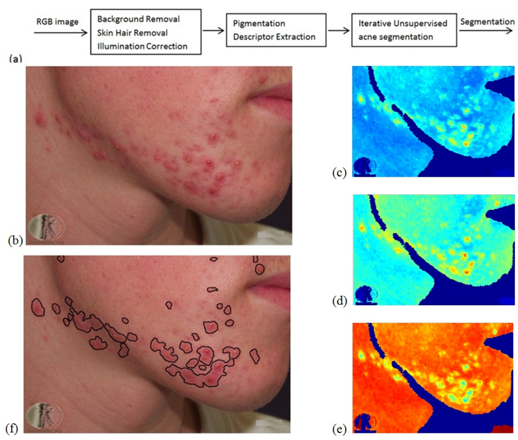


Figure 13. An initial method and result of automatic acne detection. (a) Specific steps in the acne segmentation method. (b) An example of acne disorder on human face (640×428 *pixels) from DermnetNZ (http://www.dermnetnz.org/). (c)-(e) pigmentation descriptors: hemoglobin, normalized red, and normalized green, respectively. (f) Segmentation result outlined with black borders on original image.*

COPRIN Project-Team

6. New Results

6.1. Interval analysis

6.1.1. A Contractor Based on Convex Interval Taylor

Participants: Gilles Trombettoni [correspondant], Bertrand Neveu.

Interval Taylor has been proposed in the sixties by the interval analysis community for relaxing continuous non-convex constraint systems. However, it generally produces a non-convex relaxation of the solution set. A simple way to build a convex polyhedral relaxation is to select a *corner* of the studied domain/box as expansion point of the interval Taylor form, instead of the usual midpoint. The idea has been proposed by Neumaier to produce a sharp range of a single function and by Lin and Stadtherr to handle $n \times n$ (square) systems of equations.

This paper presents an interval Newton-like operator, called `X-Newton`, that iteratively calls this interval convexification based on an endpoint interval Taylor. This general-purpose contractor uses no preconditioning and can handle any system of equality and inequality constraints. It uses Hansen's variant to compute the interval Taylor form and uses two opposite corners of the domain for every constraint.

The `X-Newton` operator can be rapidly encoded, and produces good speedups in constrained global optimization and constraint satisfaction. First experiments compare `X-Newton` with affine arithmetic[31], [19], [20]

6.2. Robotics

6.2.1. Robotics

6.2.1.1. Kinematics of wire-driven parallel robots

Participants: Laurent Blanchet, Jean-Pierre Merlet [correspondant].

The kinematics of wire robot is a complex problem because it involves both the geometrical constraints and the static equilibrium constraints as only positive tensions in the wire are possible. A major issue, that has not been addressed in the literature [16], [15], is that for a robot having n wires the forward kinematic problem (FK) (determining the possible pose(s) of the robot knowing the wire lengths, a problem that is crucial to solve for controlling the robot) cannot be solved by assuming that all n wires are under tension as the current pose of the robot may be such that only a subset of the wires may be under tension. Hence the FK problem has to be solved for **all** robots that may be derived from the initial one by removing 1 one to $n-1$ wires, each solving leading to a set of possible poses for the platform. Solving the FK for 1 wire is trivial, while for 6 wires the FK solving may be based on the already complex FK of parallel robot with rigid legs. For 2 wires it can be shown that the FK solutions can be found by solving a 12th order univariate polynomial, while for 3 wires we have shown last year by using an elimination procedure that the solutions are obtained by solving a 158th order polynomial. A very recent result of this year is that for 4 wires the order of this polynomial is 216, while no known result has been established for 5 wires (note that for 3 to 5 wires the size of the system of equations that has to be solved for the FK is larger than the one for the FK of 6-dof robot with rigid legs, a problem that has required 20 years to be solved).

Drawbacks of the elimination approach is that it does not take into account 1) that the solution should be mechanically stable, 2) that the wire tensions at the solution(s) must be positive. Hence, assuming that all solutions may be computed by the elimination approach, an a-posteriori analysis has to be performed to sort out the solutions that verify 1) and 2). We have proposed this year an efficient method to determine if a solution was mechanically stable [13]. But another major issue with the elimination method is that it leads to high order polynomial that cannot be safely numerically solved. To address this problem and 2) we are considering a numerical algorithm based on interval analysis, that consider also the tension as unknowns, hence allowing to search only for solution(s) with positive tensions [21].

Another issue for wire-driven parallel robots is the concept of redundancy. Having more wires than dof to be controlled is interesting for increasing the workspace of the robot. But it is believed that redundant wires may also be used to better distribute the load among the wires. Unfortunately we have shown for the $N-1$ robot (all N wires connected at the same point on the platform) with non elastic wires that whatever N there will be at most 3 wires under tension simultaneously [25] and consequently that tension management is not possible (with 3 wires the tensions is uniquely determined). If the wires are elastic, then tension management is possible but the positioning error is very sensitive to errors in the stiffness model [24]. Hence new method for tension management should be devised and we have explored some possibilities [23]. Still there is some magic in wire-driven parallel robots: in spite of all the uncertainties prototypes work quite well, a phenomenon which has been explained through a sensitivity analysis [24], [22].

Finally we address the management of modular robots, whose geometry can be adapted to various tasks and different objects to be manipulated, especially for very large scale robot [28], that may be used in industry for maintenance and logistics (see the Cablebot project in section 8.2.1).

6.2.1.2. Robot Calibration

Participants: Thibault Gayral, David Daney [correspondant], Jean-Pierre Merlet.

6.2.1.2.1. Experimental calibration of a high-accuracy space telescope

A collaborative work began in October 2010 with Thales Alenia Space on the calibration of the mechanical structure of a space telescope. Its architecture is based on a parallel manipulator (of the active wrist 6-PUS type, which has been patented by COPRIN) and is used to correct the relative position of two mirrors. The aim is to reach a micrometer accuracy in order to obtain a suitable quality of the images provided by the telescope. Thus, a complete model of the space telescope needs to be developed and validated through calibration. Since high velocity is not required in such an application, the dynamic effects can be neglected and only geometric and/or static calibration has to be considered. Moreover, measurements for calibration were performed in a clean room under controlled pressure, temperature and humidity conditions to minimize the influence of the non-geometric errors. Thus, two possible static inaccuracy sources were identified and modeled: one from the deformation of the mobile platform and the other resulting from the behavior of the flexure joints. Three incremental models of the flexure joints were developed and compared: a spherical joint model, a model issued from the beam theory and a stiffness model. Results of calibration using an accurate measurement system of photogrammetry showed that the flexure joints can be modeled by perfect spherical joints due to the small workspace of the telescope. Concerning the mobile platform deformation, two models were developed. Good accuracy results were obtained for both models. The developed models allowed us to explain how the model errors are directly accounted in the parameter identification during calibration. This resulted in different sets of identified parameters which all enable a good positioning accuracy. Those differences were explained and results of calibration allow a proper choice of the model of the mobile platform deformation. Considering this model, a positioning accuracy of some micrometers was finally reached after calibration with only position and orientation measurements of the mobile platform, which should allow the calibration of the telescope in space [33]. This is currently under study using interferometric measurements on the prototype of the space telescope.

6.2.1.2.2. Calibration of a cable-driven robot

To improve the accuracy of a cable manipulator, it is necessary to identify the uncertainties of its model. The cable robots, studied in the ANR funded project Cogiro (see section 8.1.1.2), are theoretically redundantly actuated: the number of powered wires is larger than the number of degrees of freedom of the manipulator (however see section 6.2.1.1 about the reality of this redundancy).

In 2011 an over-constrained prototype was self-calibrated (the identification of the parameters does not need additional external measurement), under some assumptions on the cable properties [17], [29]. We will apply our recent calibration methods on the large scale robot prototype developed for the Cogiro project at the very end of this year.

6.2.1.2.3. Cable properties

Quite often cable-driven robot analysis assume mass-less and non-elastic wires. We proposed a method based on interval analysis to judge the validity of this assumption for a particular robot in a specific workspace. Our aim is to use this method in order to determine a a region within the robot workspace for which the hypothesis is valid and consequently for which self calibration of the robot is possible. Indeed, the assumption on the cable properties is not acceptable over the full workspace of the large scale robot developed in the Cogiro project. Still a self-calibration is possible if calibration poses are chosen within a specific subpart of the workspace. A more efficient calibration approach is in progress with additional measures and a more complex model (static and elasticity). The results has been published in [18], [30].

6.2.1.3. Assistance robotics

Participants: David Daney, Claire Dune, Jean-Pierre Merlet [correspondant], Yves Papegay, Odile Pourtallier.

As mentioned earlier in the report we have started in 2008 a long term strategic move toward assistance robotics, with the objectives of providing low-cost, simple to control, robotized smart devices that may help disabled, elderly and handicapped people in their personal life, provide also assistance to family and caregivers while allowing doctors to get better and objective information on the health state of the end-user. Our credo is that these devices have to be adapted to the end-user and to its everyday environment (by contrast with the existing trend of focusing on a "universal" robot, to which the end-user and its environment have to adapt). As for cost reasons we intend to use only standard hardware adaptation has to be taken into account at the very early stage of the system design and uncertainties in the physical instances of our systems are also to be considered.

For validation purposes we have developed a flat in order to explore various full scale scenarii that cover a part of the daily life of an elderly, to develop specific assistance devices and to test them [1]. Our activity in this field is concentrated on transfer, manipulation, walking monitoring, rehabilitation and the use of virtual reality. We are also investigating how such complex environments with multiples smart agents, quite heterogeneous from a computing viewpoint, but that have to cooperate, may be programmed. All these topics are in accordance with the one of the large scale initiative PAL [2] of which we are an active member.

6.2.1.3.1. Transfer and manipulation

Participants: François Chaumette [Lagadic], Jean-Pierre Merlet, Rémy Ramadour.

For transfer operation we are using the `MARIONET-ASSIST` robot (see section 6.2.1.4) that is installed in our flat. Currently we use 4 winches with the wires connected at the same point on the platform, hence providing 3 translational degrees of freedom. This low-cost, low-intrusivity robot has proved to be very effective for transfer operation. Apart of transfer operation robot may be used for manipulation. Adding one or several low-cost cameras (the cost being here a fundamental constraint), visual-servoing control is used to provide a whole new set of useful services such as grasping objects in order to bring them to the end-user (if they are too heavy, too far, high or low), or cleaning the table after lunch. Using a parallel crane robot, we are able to cover a large workspace, the vision-control allowing us to obtain the precision required by the manipulation of daily-life objects. The collaborative implementation of the vision and the kinematic control of the robot gives us a way to make best use of the advantages of both parts, while overcoming their respective drawbacks.

Given a region where the object of interest belongs, the first step is to detect it in an evolutive environment. A segmentation is made, robust to luminance variations and perspective projections. The vision is then used to move the platform toward a desired position relatively to the target. In order to execute this task, some carefully chosen features are measured, allowing to estimate the incremental displacement required to move the end-effector to the desired place. We use the library ViSP for both the detection and the visual-control part.

[1] pictures of this assistive flat are available at http://www-sop.inria.fr/coprin/prototypes/main.html

[2] http://pal.inria.fr

Experimental results were obtained using a platform with 3 degrees of freedom and a single camera, grasping a single object and moving it from a place to another. We used for that basic image data such as 2D moments, allowing a fast computing and yet robustness in measurements. We currently address to generalize this manipulation to other configurations and to evaluate its robustness to calibration errors and other uncertainty sources. We also are looking for a global paradigm merging both the vision-based kinematic model and the mechanical one, which could significantly improve the efficiency of the experiment, while reducing the mathematical complexity behind each kinematic model considered on its own.

6.2.1.3.2. Walking monitoring

Participants: Claire Dune [Handibio], Jean-Pierre Merlet.

We use the walking aids `ANG-light` and `ANG-II` (see section 6.2.1.4) to monitor the trajectory of the walking aid. The on-board sensors of these aids allow to evaluate the step pattern, gait asymmetry,...during daily walking, hence providing an health monitoring system that is always available. `ANG-light` has been tested last year with 24 subjects that were themselves instrumented (accelerometers in the wrists and knees, pressure sensors in the shoes) and were asked to perform two different trajectories twice with/without the walking aid. The purposes were:

- to determine pertinent walking indicators
- to obtain a “gold” standard of these indicators for non pathological walking, taking into account the normal variability of the walking pattern
- to determine if indicators obtained with the walking aid may led to an accurate estimation of the indicators when the walking aid is not used

Several indicators have been determined after the analysis of these data. In a second phase the inclusion test of elderly people (30 subjects) is taking place at the CHU of Nice-Cimiez and will last until the first trimester of 2013. An analysis of the collected data, in close collaboration with the doctors, will allow to determine if the proposed walking indicators are pertinent.

Another interest of the walking aids `ANG` is that they allow to collect significant information for mobility during their daily use: slope and surface quality of the sidewalks and automatic detection of lowered kerbs with a ranking of their convenience. It will be interesting for a community to share such information that is collected by the community members. For that purpose we propose to use collective maps, such as OpenStreeMap, which allow for map annotation. To validate this concept we have used `ANG-light` to automatically annotate the map of the Inria Sophia site with pertinent information for walking aid and wheelchair users [3].

6.2.1.3.3. Rehabilitation

Participants: David Daney, Mandar Harshe, Sami Bennour, Jean-Pierre Merlet [correspondant].

The focus of our work is on analyzing knee joint motion during a walking activity. The main principle of the system is to observe relative motions of the collars attached to tibia and femur. The measurement of the motion of these collars is based on the wire actuated parallel robot architecture (using the `MARIONET-REHAB` robot, see section 6.2.1.4). To increase the reliability of our analysis, and decrease the influence of Skin Tissue Artifacts (STA), we also incorporate a passive wire measurement system, IR camera based motion capture system, accelerometers, and force sensors to measure human motions.

Measurements in the global frame and collar specific local frames give precise data to reconstruct collar (and thus, knee joint) motion. The system developed already incorporates the optical motion capture, inertial measurement units and the wire sensors for comprehensive coordinated measurements of the motion of the knee.We have performed preliminary trials on three subjects for walking motion.

In the past year, we worked on processing the data to obtain pose and orientation information of knee joint. Data obtained from the trials was analyzed and post-processing steps were implemented to reduce noise and errors. In order to perform sensor fusion, we implemented a probabilistic estimation based method to estimate the pose.

[3] see http://www-sop.inria.fr/coprin/prototypes/main.html#ang

The results from these analysis have allowed us to identify the merits of our approach and also helped us identify improvements that are needed. We have also identified the possible changes to our mathematical model that could allow use of interval analysis tools along with probabilistic estimation methods. We have identified the changes needed to the hardware setup that will help reduce the sensor noise and error. These changes once implemented will allow us to improve the usability of the system and also point us towards newer areas for further investigation, including, for example, effect of sensor placement, collar design, and interval based extended Kalman Filters for pose estimation[8], [9].

6.2.1.3.4. Virtual reality

Virtual reality has proved to be an effective mean for dealing with rehabilitation, provided that motion is added to the 3D visual feedback. The `MARIONET-VR` robot, together with our motion base (see section 6.2.1.4) may provide very realistic motion in an immersive room and we will start in 2013 a collaboration with the VR4I and REVES project-teams on this issue. A first task will be to install a moving walking treadmill in the immersive room at Sophia and to combine motion of the treadmill with 3D viewing.

6.2.1.3.5. Programming

In our opinion there will not be a single assistance device that will be able to offer all the required help that may be needed but rather numerous redundant smart agents that are able to perform very efficiently (at low cost and with a small energy consumption) a set of specific tasks. Such agents must be able to communicate and possibly will have to cooperate in some cases (e.g. after a fall). They will be heterogeneous from a computer view point as the used agents will change according to technological advances and to the trajectory of life of the end-users. If the manual programming of a single agent is possible (although quite complex for some of them) the overall system cannot be managed in that way: we need an unifying framework for this development. For that purpose we are currently investigating the use of HOP, a multi-tier programming language for the diffuse Web developed in the INDES project-team. Already one of our wire-driven parallel robot `MARIONET-SCHOOL` has been programmed in HOP and can be seen as a web resource.

6.2.1.4. Prototypes

Participants: Julien Alexandre Dit Sandretto, David Daney, Claire Dune, Jean-Pierre Merlet [correspondant].

Experimental works are a key point in the field of service robotics: it allows for validating concepts, getting feedback from the end-users and discovering new problems. We have extensively developed prototypes this year [4]

6.2.1.4.1. Wire-driven parallel robots

The `MARIONET` family is now constituted of

- `MARIONET-REHAB` (2004-): using up to 7 linear actuators it is mainly used for rehabilitation and health monitoring although it is also a very fast pick-and place robot
- `MARIONET-CRANE` (2008-): a very large 6-dof rescue crane, portable and autonomous that can be deployed in 10 minutes and has a lifting capability of 2 tons
- `MARIONET-ASSIST` (2011-): a robot deployed in our flat with up to 6 winches, that is used for transfer operation and health monitoring, with a lifting ability of about 2000 kg
- `MARIONET-VR` (2011-): using up to 6 linear actuators it will be used in the Sophia immersive room for simulation and rehabilitation. It allows to fully lift a person and has been physically installed in the immersive room this year although it is not fully operational
- `MARIONET-SCHOOL` (2012-): a set of very low-cost robots that are intended to be used for dissemination. They fit in a small suitcase and can be deployed on a table or over a full classroom. We believe that such robots may be used by roboticists but also by researchers from other domains that may use the motion of the robot(s) to illustrate scientific concepts. Currently we have 3 such robots (one using Lego components, one with step motors and one with servos)

[4] pictures and videos of our prototypes are available at http://www-sop.inria.fr/coprin/prototypes/main.html or on our YouTube channel http://www.youtube.com/user/CoprinTeam/videos?flow=grid&view=0

6.2.1.4.2. Walking aids

The *Assistive Navigation Guide* (ANG) family is based on commercially available Rollators. with several objectives (we mention only a few of them):

- fall prevention/detection: fall is a major problem for elderly
- mobility help: provide an on-demand mobility help
- gait pattern monitoring: we believe that being able to monitor the trajectory of the walking aid will provide useful information on the gait pattern of the user

For reaching these objectives we have developed two walking aids:

- `ANG-light`: a walking aid with encoders in the wheels, 3D accelerometer, gyrometer and GPS. These sensors allow to measure the trajectory of the walking aid and several features of the user's gait. This walking aid is currently being used at CHU Nice, see section 6.2.1.3.2 . A replica of `ANG-light` is currently under development at the Handibio laboratory of Toulon and will include force sensors in the handles to get measurement of muscular activities while walking.
- `ANG-II`: this aid is an evolution of the motorized walker `ANG`, with a lower weight and better integration

6.2.1.4.3. Other devices

As seen in the previous sections we have focused our work on mobility as it has been identified as a major priority during our two years interview period. Another priority is fall management: it is adressed with the `ANG`'s) but requires that the patient uses a walking aid. To obtain a better coverage we have developed an instrumented vest that includes an Arduino Lilypad microcontroller to monitor a 3D accelerometer and verticality sensors in order to detect a fall. This system may communicate an alarm through wifi, zigbee or even infrared signals. Apart of its low cost the system is washable and has a very low power consumption (we are considering energy harvesting to further increase its energy autonomy).

Two other important issues for assistance robotics is activities monitoring and patient localization. In cooperation with the STARS project-team an activity monitoring system within the Dem@care project. Its purpose is to log specif events such as taking a pen, setting on a kettle, ...using only simple sensors such as proximity and distance sensors, switches, that are more reliable and less complex than using a vision system. We have provided to the CHU a system that allows to monitor up to 208 different events. In the same manner we are working on a localization system of elderly in a room based only on distance sensors, that will be less intrusive and/or may complement the measurements of a vision system.

Rehabilitation is also part of our activities. Apart of the `MARIONET-REHAB` system we plan to investigate the use of Sophia immersive room for that purpose. To complement the motion that will be provided by the `MARIONET-VR` robot we have bought a 6-dof motion base from Servos with a nominal load of 150kg, that we have modified to accommodate our needs. We will use also two lifting columns with a load of 100 kg each, that will allow to manage motion for rehabilitation apparatus such as treadmill.

6.3. Miscellaneous results

6.3.1. Symbolic tools for modeling and simulation

Participant: Yves Papegay.

This activity is the main part of a long-term ongoing collaboration with Airbus whose goal is to directly translate the conceptual work of aeronautics engineers into digital simulators to accelerate aircraft design.

An extensive modeling and simulation platform has been designed which includes a dedicated modeling language for the description of aircraft dynamics models in term of formulae and algorithms, and a symbolic compiler producing as target an efficient numerical simulation code ready to be plugged into a flight simulator, as well as a formatted documentation compliant with industrial requirements of corporate memory [10].

Implementation of this platform is a modeling and simulation environment based on symbolic computation tools. It contains several components :

- a model editor, that makes it possible and easy to enter the whole set of equations describing large and complex industrial models,
- an highly interactive and modular evaluation workbench allowing to simulate the models and to visualize the results inside the modeling environment with the benefits for the designer of being able to directly use all its computational functionnalities.
- a C code generator which, using these models, automatically generates the numerical real-time simulation engines
- a technical documentation generator

To finalize the transfer of the technology demonstrated by our prototype to our industrial partner, an extensive collection of testing and corresponding improvements have been done in 2012. This step has delayed the delivery of the final version of our modeling and simulation environment to Airbus until November 2012.

6.3.2. Multi-agent aircraft design

Participant: Yves Papegay.

The modeling environment described in the previous section is used, in collaboration with other teams at Airbus, in the framework of the ID4CS project founded by ANR and dedicated to multi-agent optimization of large scale system.

Several models of aircraft engines and of aircraft have been developed as user cases for the project.

In 2012 we focused on automatic generation of agent code based on models that is ready for integration in the ID4CS platform prototype.

6.3.2.1. Equilibrium strategies for linked Electricity and CO2 markets

Participant: Odile Pourtallier.

In collaboration with M. Bossy (Inria -TOSCA Team) and N. Maïzi (CMA - Mines Paristech) O. Pourtallier has pursued the study of equilibrium model for coupled electricity and CO2 allowance exchange markets (see also Section 7.3). We have mainly focused on the determination of Nash equilibrium for the coupled electricity and carbon markets with the assumption that the producers maximize their market shares. Nash equilibrium have been obtained by using as an intermediary step a decreasing auction mechanism.

We have also pursued an indifference pricing methodology which is presented in more details in Inria -TOSCA Team section.

DAHU Project-Team

5. New Results

5.1. Distributed data management

Participants: Serge Abiteboul, Émilien Antoine, Cristina Sirangelo, Nadime Francis, Luc Segoufin.

Distributed knowledge base. We are developing the system Webdamlog [16], [13], [14] to address the challenges faced by everyday Web users, who interact with inherently heterogeneous and distributed information. Managing such data is currently beyond the skills of casual users. In Webdamlog, we see the Web as a knowledge base consisting of distributed logical facts and rules. The objective is to enable automated reasoning over this knowledge base, ultimately improving the quality of service and of data. The system supports the Webdamlog language, a Datalog style language with rule delegation.

Deduction in uncertain worlds. Motivated by reasoning in distributed environments in which disagreements arise between different actors, we study in [17] deduction (captured by datalog programs) in the presence of inconsistencies (induced by functional dependency (FD) violations). We adopt an operational semantics for datalog with FDs based on inferring facts one at a time, while never violating the FDs. This yields a set of possible worlds that we capture by c-tables of possibly exponential size. We propose to use probabilities to measure this nondeterminism and define a probabilistic semantics that can be captured by probabilistic conditional tables. Not surprisingly, we show that computing the probability of a query answer in our setting is expensive, which leads us to introduce a sampling algorithm to estimate answer probabilities. We then turn our attention to the problem of explaining why a particular answer holds. This leads us to consider two novel notions: the most influential extensional facts, and the most likely proofs for an answer. We study algorithms for ranking facts and proofs based on their contribution to the derivation of an answer. Finally, we consider how our framework can be adapted to a distributed setting, and in particular, how sampling can be performed in a distributed manner.

Access rights in a distributed setting. We started considering access right issues in Webdamlog. This is related to specifying access right on views in standard databases. There is also the issues of controlling rules that are run locally but were specified by other peers.

Incomplete information in Web data. Incomplete information often arises from the integration of different Web data sources, as well as from the exchange of data between communicating Web applications. The semantics of incompleteness (i.e. which possible complete databases are represented by an incomplete one) depends on the context and the particular scenario where incompleteness raises from. We have studied how to deal with the presence of incomplete information under different possible semantics. We have in particular studied in which condition it is possible to query incomplete data “naively", i.e. as if it were complete. We have exhibited “natural" fragments of first order logic for which naive evaluation is possible, under different semantics.

Graph data management. Graph structured data can be found in new emerging applications such as RDF and linked data, or social networks. The peculiarity of queries over graphs is that they are interested in both data carried by the graph and in the graph topology; they are often based on reachability patterns. In a distributed setting it is very common to be able to query only a partial description or a "view" of the graph. We studied the problem of answering queries using only the information provided by the views. The presence of a form of recursion in views and queries presents new challenges. We found restricted classes of graph views and queries that allow efficient query answering over views.

5.2. Tree automata theory

Participants: Luc Segoufin, Serge Abiteboul, M Praveen.

Tree automata We studied the expressive power of a subclass of regular tree languages. We gave a decidable characterization of those languages that are "piecewise testable", i.e. definable using boolean combination of existential first-order formulas [12].

Automata with counters. We studied extending techniques used in standard Petri nets to other models. We extended the Rackoff technique to decide coverability and boundedness problems for Strongly Increasing Affine nets, a subclass of Affine nets [20].

Languages on trees. We studied in [18] highly expressive query languages for unordered data trees, using as formal vehicles Active XML and extensions of languages in the while family. All languages may be seen as adding some form of control on top of a set of basic pattern queries. The results highlight the impact and interplay of different factors: the expressive power of basic queries, the embedding of computation into data (as in Active XML), and the use of deterministic vs. nondeterministic control. All languages are Turing complete, but not necessarily query complete in the sense of Chandra and Harel. Indeed, we show that some combinations of features yield serious limitations, analogous to FOk definability in the relational context. On the other hand, the limitations come with benefits such as the existence of powerful normal forms. Other languages are "almost" complete, but fall short because of subtle limitations reminiscent of the copy elimination problem in object databases.

Probabilistic XML. In [15], we study the problem of, given a corpus of XML documents and its schema, finding an optimal (generative) probabilistic model, where optimality here means maximizing the like- lihood of the particular corpus to be generated. Focusing first on the structure of documents, we present an efficient algorithm for finding the best generative probabilistic model, in the absence of constraints. We further study the problem in the presence of integrity constraints, namely key, inclusion, and domain constraints. We study in this case two different kinds of generators. First, we consider a continuation-test generator that performs, while generating documents, tests of schema satisfiability ; these tests prevent from generating a document violating the constraints but, as we will see, they are computationally expensive. We also study a restart generator that may generate an invalid document and, when this is the case, restarts and tries again. Finally, we consider the injection of data values into the structure, to obtain a full XML document. We study different approaches for generating these values.

Infinite alphabet. We studied the complexity of satisfiability of linear temporal logics extended to reason about repetitions of values from an infinite data domain. We refined an existing result that reduced this problem to Petri net reachability, and showed that it can be reduced to the coverability problem. Using this refinement, we gave the precise complexity of the satisfiability problem. We also characterized the complexity of satisfiability for many fragments and extensions of the logic.

DREAM Project-Team

6. New Results

6.1. Diagnosis of large scale discrete event systems

Participants: Marie-Odile Cordier, Sophie Robin, Laurence Rozé, Yulong Zhao.

The problem we deal with is monitoring complex and large discrete-event systems (DES) such as an orchestration of web services or a fleet of mobile phones. Two approaches have been studied in our research group. The first one consists in representing the system model as a discrete-event system by an automaton. In this case, the diagnostic task consists in determining the trajectories (a sequence of states and events) compatible with the sequence of observations. From these trajectories, it is then easy to determine (identify and localize) the possible faults. In the second approach, the model consists in a set of predefined characteristic patterns. We use temporal patterns, called chronicles, represented by a set of temporally constrained events. The diagnostic task consists in recognizing these patterns by analyzing the flow of observed events.

6.1.1. Distributed monitoring with chronicles - Interleaving diagnosis and repair - Making web services more adaptive

Our work addresses the problem of maintaining the quality of service (QoS) of an orchestration of Web services (WS), which can be affected by exogenous events (i.e., faults). The main challenge in dealing with this problem is that typically the service where a failure is detected is not the one where a fault has occurred: faults have cascade effects on the whole orchestration of services. We have proposed a novel methodology to treat the problem that is not based on Web service (re)composition, but on an adaptive re-execution of the original orchestration. The re-execution process is driven by an orchestrator Manager that takes advantage of an abstract representation of the whole orchestration and may call a diagnostic module to localize the source of the detected failure. It is in charge of deciding the service activities whose results can be reused and may be skipped, and those that must be re-executed.

This year, we have improved the prototype and worked on a journal paper that will be submitted in 2013.

6.1.2. Scenario patterns for exploring qualitative ecosystems

This work aims at giving means of exploring complex systems, in our case ecosystems. We proposed to transform environmental questions about future evolution of ecosystems into formalized queries that can be submitted to a simulation model. The system behavior is represented as a discret event system described by a set of interacting timed automata, the global model corresponding to their composition on shared events. To query the model, we have defined high-level generic query patterns associated to the most usual types of request scenarios. These patterns are then translated into temporal logic formula. The answer is computed thanks to model-checking techniques that are efficient for analysing large-scale systems. Five generic patterns have been defined using TCTL (Timed Computation Tree Logic) “WhichStates”, “WhichDate”, “Stability”, “Always”, “Safety”. Three of them have been implemented using the model-checker UPPAAL.

The approach has been experimented on a marine ecosystem under fishing pressure. The model describes the trophodynamic interactions between fish trophic groups as well as interactions with the fishery activities and with an environmental context. A paper has been accepted for publication in the Environmental Modelling Software Journal [52].

6.1.3. Controler synthesis for dealing with “How to” queries

We extended the approach to deal with “How to” queries. As before, we rely on a qualitative model in the form of timed automata and on model-checking tools to answer queries. We proposed and compared two approaches to answer questions such as “How to avoid a given situation ?”(safety query). The first one exploits controller synthesis and the second one is a “generate and test” approach. We evaluated these two approaches in the context of an application that motivates this work, i.e the management of a marine ecosystem and the evaluation of fishery management policies. The results have been accepted for publication in [17].

More recently, we use similar methodological tools to model herd management on a catchment and analyse the best/optimal farming practices in order to reduce nitrate pollution due to livestock effluents. An hybrid model has been built using hierarchical timed automata. Scenarios can already be simulated and evaluated. We currently work on adapting controler synthesis tools in order to get the best strategies. This work is made in collaboration with our colleagues of INRA.

6.2. Machine learning for model acquisition

Participants: Marie-Odile Cordier, Thomas Guyet, Simon Malinowski, René Quiniou, Sid Ahmed Benabderrahmane.

Model acquisition is an important issue for model-based diagnosis, especially while modeling dynamic systems. We investigate machine learning methods for temporal data recorded by sensors or spatial data resulting from simulation processes. Our main objective is to extract knowledge, especially sequential and temporal patterns or prediction rules, from static or dynamic data (data streams). We are particularly interested in mining temporal patterns with numerical information and in incremental mining from sequences recorded by sensors.

6.2.1. Mining temporal patterns with numerical information

We are interested in mining interval-based temporal patterns from event sequences where each event is associated with a type and time interval. Temporal patterns are sets of constrained interval-based events. This year we have begun to work on multiscale temporal abstraction to represent time series by codewords at different temporal and amplitude scales. We have improved the method of Wang et al. [70] by introducing Dynamic Time Warping to compute better codewords for time series abstraction. The codeword-based time series representation is then used by QTIPrefixSpan [3] to extract temporal patterns. A paper is in preparation. We are also working on a multivariate version of the method for mining multivariate temporal patterns at different resolution levels.

6.2.2. Incremental sequential mining

Sequential pattern mining algorithms operating on data streams generally compile a summary of the data seen so far from which they compute the set of actual sequential patterns. We propose another solution where the set of actual sequential patterns are incrementally updated as soon as new data arrive on the input stream. Our work stands in the framework of mining an infinite unique sequence. Though being of great importance, this problem has not received a similar attention as mining from a transaction database. Our method [13] provides an algorithm that maintains a tree representation (inspired by the PSP algorithm [56]) of frequent sequential patterns and their minimal occurrences [54] in a window that slides along the input data stream. It makes use of two operations: deletion of the itemset at the beginning of the window (obsolete data) and addition of an itemset at the end of the window (new data). The experiments were conducted on simulated data and on real data of instantaneous power consumption. The results show that our incremental algorithm significantly improves the computation time compared to a non-incremental approach [14].

6.2.3. Incremental learning of preventive rules

The problem is to learn preventive rules in order to avoid malfunctioning on smartphones. A monitoring module is embedded on the phones and sends reports to a server. Reports are labeled with a normal or abnormal label. From this set of reports new rules are learned. As a lot of smartphones are supervised, it is impossible to store all the reports. Therefore incremental learning has to be used.

Last year, we achieved two main tasks: a report database has been built in order to test the future algorithms, and a new algorithm [20] has been developed for implementing an incremental version of the algorithm AQ21 [72].

6.2.4. Multiscale segmentation of satellite image time series

Satellite images allow the acquisition of large-scale ground vegetation. Images are available along several years with a high acquisition frequency (1 image every two weeks). Such data are called satellite image time series (SITS). In [12] , we present a method to segment an image through the characterization of the evolution of a vegetation index (NDVI) on two scales: annual and multi-year. We test this method to segment Senegal SITS and compare our method to a direct classification of time series. The results show that our method using two time scales better differentiates regions in the median zone of Senegal and locates fine interesting areas (cities, forests, agricultural areas).

6.2.5. Mining a big unique graph for spatial pattern extraction

Researchers in agro-environment needs a great variety of landscapes to test the agro-ecological models of their scientific hypotheses. As the representation of real landcapes necessitates lots of on-land measures, good big representations are difficult to acquire. Working with landscape simulations is then an alternative to get a sufficient variety of experimental landscapes. We propose to extract spatial patterns from a well described geographic area and to use these patterns to generate realistic landscapes. We have begun the exploration of graph mining techniques to discover the relevant spatial patterns present in a graph expressing the spatial relationships between the agricultural plots as well as the roads, the rivers, the buildings, etc., of a specific geographic area.

This year, we have been working on extending algorithm gSPAN [73] with an adaptive support threshold and with a taxonomy to be able to extract interesting patterns involving agricultural plots with rare features. We plan to submit a paper in 2013.

6.3. Decision aiding with models and simulation data

Participants: Louis Bonneau de Beaufort, Tassadit Bouadi, Marie-Odile Cordier, Véronique Masson, Florimond Ployette, René Quiniou, Karima Sedki.

Models can be very useful for decision aiding as they can be used to play different plausible scenarios for generating the data representing future states of the modeled process. However, the volume of simulation data may be very huge. Thus, efficient tools must be investigated in order to store the simulation data, to focus on relevant parts of the data and to extract interesting knowledge from these data.

6.3.1. Exploring models thanks to scenarios: a generic framework

In the framework of the Appeau project (see 7.2.1) a paper describing a generic framework for scenario exercises using models applied to water-resource management, has been written in cooperation with all the partners and published in Environmental Modelling and Software [5].

6.3.2. A datawarehouse for simulation data

The ACASSYA project 7.2.2 aims at providing experts or stakeholders or farmers with a tool to evaluate the impact of agricultural practices on water quality. As the simulations of the deep model TNT2 are time-consuming and generate huge data, we have proposed to store these simulation results in a datawarehouse and to extract relevant information, such as prediction rules, from the stored data. We have devised a general architecture for agro-environmental data on top of the framework Pentaho.

This year we have been working on the efficient computation of OLAP queries related to realistic scenarios proposed by experts in the domain. Precisely, we have devised indexing schemes to access the data in the OLAP cube. We have also worked on the visualization by a GIS (Geographical Information System) of the query results on maps of the geographical area under interest. A paper will be submitted to the COMPAG Journal in beginning 2013.

6.3.3. Efficient computation of skyline queries in an interactive context

Skyline queries retrieve from a database the objects that maximizes some criteria, related to user preferences for example, or objects that are the best compromises satisfying these criteria. When data are in huge volumes, such objects may shed light on interesting parts of the dataset. However, computing the skylines (i.e. retrieving the skyline points) may be time consuming because of many dominance tests. This is, especially the case in an interactive setting such as querying a data cube in the context of a datawarehouse.

This year we have worked at improving the formal setting of the partial materialization of skyline queries when dynamic preferences are refined online by the user. We have explicited which parts of the skyline evolve (which point are added or removed) when a new dimension is introduced in the computation. This led to an efficient incremental method for the online computation of the skyline corresponding to new user preferences [9]. An extended version of this paper is under submission to the Journal "Transactions on Large Scale Data and Knowledge Centered Systems" (TLDKS).

We are working now on a hierachical extension of our method that could be introduced in a datawarehouse context.

6.3.4. Influence Diagrams for Multi-Criteria Decision

For multi-criteria decision-making problems, we propose in [7] a model based on influence diagrams able to handle uncertainty, represent interdependencies among the different decision variables and facilitate communication between the decision-maker and the analyst. The model makes it possible to take into account the alternatives described by an attribute set, the decision-maker's characteristics and preferences, and other information (e.g., internal or external factors) that influence the decision. Modeling the decision problem in terms of influence diagrams requires a lot of work to gather expert knowledge. However, once the model is built, it can be easily and efficiently used for different instances of the decision problem. In fact, using our model simply requires entering some basic information, such as the values of internal or external factors and the decision-maker's characteristics.

6.3.5. Modeling influence propagation by Bayesian causal maps

The goal of this project is modeling shellfish fishing to assess the impact of management pollution scenarios on the *Rade de Brest*. Cognitive maps were built from interviews with fishermen. To represent and reason about these cognitive maps, we propose to use Bayesian Causal Maps making use of fishermen knowledge, particularly to perform influence propagation [11].

However, this model does not take into account the variety of influences asserted by the fishermen, but only the "mean" causal map. This year we have been working on an approach that could combine individual knowledge with belief functions in the way of Philippe Smets's Transferable Belief Model [67].

This work is done in the framework of the RADE2BREST project, involving Agrocampus Ouest and CNRS (GEOMER/LETG), funded by "Ministère de l'Ecologie" (This project is not mentioned in section 7.2 because DREAM is not an official partner of this project.).

6.3.6. Mining simulation data by rule induction

In the framework of the SACADEAU project (see 7.2.1), a paper dedicated to mining simulation data by rule induction has been published in the COMPAG Journal [8]. Both qualitative and quantitative predictions from a model of an agro-environmental system are analysed. Two approaches in rule learning from spatial data (ILP and attribute-value approaches) are compared and show that results help identify factors with strong influence on herbicide stream-pollution.

We have also participated in a collaboration for modeling the effects crop rotations the results of which were published in the Science of the Total Environment Journal [6].

6.4. Diagnostic and causal reasoning

Participants: Philippe Besnard, Louis Bonneau de Beaufort, Marie-Odile Cordier, Yves Moinard, Karima Sedki.

Stemming on [29], [30], [31], [32], [33], we have designed an inference system based on causal statements. This is related to diagnosis (observed symptoms explained by faults). The aim is to produce possible explanations for some observed facts. Previously existing proposals were ad-hoc or, as in [36], [47], they were too close to standard logic in order to make a satisfactory diagnosis. A key issue for this kind of work is to distinguish logical implication from causal links and from ontological links. This is done by introducing a simple causal operator, and an *is-A* hierarchy. These two operators are added to a restricted first order logic of the Datalog kind (no function symbols). Then, our system produces elementary *explanations* for some set of observed facts. Each explanation links some facts to the considered observation, together with a set of atoms called the *justifications*: The observation is explained from these facts, provided the justifications are possible (not contradicted by the available data). This formalism has also been translated into answer set programming [57], [58]).

This year, we have extended our formalism in order to deal with more complex problems such as finding explanations for the hurricane Xynthia (2010, February 28). In such situations, there are many data and many possible elementary explanations can be examined. This involves an extension of our formalism, in order to deal with more complex chains of causations and *is-A* links. We are on the way to end this task. Our formalism makes precise what all these possible explanations are. Then, in order to deal with so many possible complex explanations, we integrate this causal formalism into an argumentation framework. Logic-based formalizations of argumentation [34] take pros and cons for some conclusion into account. These formalizations assume a set of formulae and then exhaustively lay out arguments and counterarguments. This involves providing an initiating argument for the inference and then providing undercuts to this argument, and then undercuts to undercuts. So here our causal formalism provides a (rather large) set of explanations, and the argumentation part allows to select the best ones, under various criteria.

Then, since answer set programming can easily deal with logical formalisms, the argumentation part will be incorporated into our already existing answers set programming translation of the causal formalism. Regarding answer set programming, we have also examined some more difficult examples [16] and participated to a chapter in the to be published "Panorama de l'intelligence artificielle. Ses bases méthodologiques, ses développements" [19].

E-MOTION Project-Team

5. New Results

5.1. Perception and Situation Awareness in Dynamic Environments

5.1.1. Sensor Fusion for state parameters identification

Participants: Agostino Martinelli, Chiara Troiani.

5.1.1.1. Problem adressed and background

The general framework based on the new concept of continuous symmetry developed during the last two years (see [67] for a detailed description of this framework) has been extensively applied to investigate the visual inertial structure from motion problem. This problem was already considered in 2011. During 2012 more general results have been found. Special attention has been devoted to identify the conditions under which the problem has a finite number of solutions. Specifically, it has been shown that the problem can have a unique solution, two distinct solutions and infinite solutions depending on the trajectory, on the number of point-features and on their layout and on the number of camera images. The investigation has also performed in the case when the inertial data are biased, showing that, in this latter case, more images and more restrictive conditions on the trajectory are required for the problem resolvability.

5.1.1.2. Theorical results

The new results have been published on the journal of Transaction on Robotics [68], in a technical report [43] and submitted to the International Journal of Computer Vision. We have also considered the case of structured light. Specifically, we have considered a sensor assembling (from now on *aerial vehicle*) consisting of a monocular camera and inertial sensors. Additionally, a laser pointer is mounted on the aerial vehicle and it produces a laser spot. The laser spot is observed by the monocular camera and it is the unique point feature used in the proposed approach. We focus our attention to the case when the aerial vehicle moves in proximity of a planar surface and in particular when the laser spot belongs to this surface. We introduced two novel contributions. The former is the analytical derivation of all the observable modes, i.e., all the physical quantities that can be determined by only using the inertial data and the camera observations of the laser spot during a short time-interval. This derivation was based on the framework introduced in [67]. Specifically, it is shown that the observable modes are: the distance of the vehicle from the planar surface; the component of the vehicle speed, which is orthogonal to the planar surface; the relative orientation of the vehicle with respect to the planar surface; the orientation of the planar surface with respect to the gravity. The second contribution is the introduction of a simple recursive method to perform the estimation of all the aforementioned observable modes. This method is based on a local decomposition of the original system, which separates the observable modes from the rest of the system. The method has been validated by using synthetic data. Additionally, preliminary tests with real data are provided and more complete experiments are in progress. The presented approach can be integrated in the framework of autonomous take-off and landing, safe touch-down and low altitude manoeuvres even in dark or featureless environment. These results have been published in the iros conference [31]

5.1.1.3. Experimental results

In parallel to this theoretical activity an experimental activity has been carried out in order to deploy our technologies to industrial partners. To this regard, we had a collaboration with the company Delta Drone in Grenoble. In this framework we introduced a new method to localize a micro aerial vehicle (MAV) in GPS denied environments and without the usage of any known pattern. The method exploits the planar ground assumption and only uses the data provided by a monocular camera and an inertial measurement unit. It is based on a closed solution which provides the vehicle pose from a single camera image, once the roll and the pitch angles are obtained by the inertial measurements. Specifically, the vehicle position and attitude can uniquely be determined by having two point features. However, the precision is significantly improved by using three point features. The closed form solution makes the method very simple in terms of computational cost and therefore very suitable for real time implementation. Additionally, because of this closed solution, the method does not need any initialization. We have implemented this method on the platform available in our lab. This is a *Pelican* from *Ascending Technologies* equipped with an Intel Atom processor board (*1.6 GHz, 1 GB RAM*) (figure 1).

Figure 1. AscTec Pelican quadcopter equipped with a monocular camera.

Our sensor suite consists of an Inertial Measurement Unit (*3-Axis Gyro, 3-Axis Accelerometer*) belonging to the Flight Control Unit (FCU) "AscTec Autopilot" , and a monocular camera (*Matrix Vision mvBlueFOX*, FOV: 130 deg). The camera is calibrated using the Camera Calibration Toolbox for Matlab by J.Y. Bouguet at caltech. The calibration between IMU and camera has been performed using the Inertial Measurement Unit and Camera Calibration Toolbox in [66]. The IMU provides measurements update at a rate of $100Hz$, while the camera framerate is $10Hz$. The Low Level Processor (LLP) of our Pelican is flashed with the *2012 LLP Firmware* and performs attitude data fusion and attitude control. We flashed the High Level Processor (HLP) with the asctec_hl_firmware [48]. The onboard computer runs linux 10.04 and ROS (Robot Operating System). We implemented our method using ROS as a middleware for communication and monitoring . The HLP communicates with the onboard computer through a FCU-ROS node. The communication between the camera and the onboard computer is achieved by a ROS node as well. The presented algorithms are running online and onboard at $10Hz$.

The scenario setup is shown in figure 3 . Since our lab is not yet equipped with a Motion Capture System, we used an ARToolKit Marker with the only aim of having a ground truth to evaluate the performance of our approach. The estimation of the camera pose provided by the marker is not used to perform the estimation. The marker is positioned such that it's reference frame is coincident with the configuration shown in figure 3 . The three features considered are the center of the three little balls in figure 3 . The use of three blob markers instead of natural features is only related to the need to get a ground truth. The information related to the pattern composed by the 3 features ($D = 0.25m$, $\gamma_1 = 60deg$, $\gamma 2 = 120deg$) is only used to evaluate the performance of our approach. The algorithm does not require any information about the features configuration.

Figure 4 and 5 show respectively the position and the attitude by using the proposed approach. The estimated values are compared with the ground truth obtained with the ARToolkit marker. From figure 4 we see that the difference between our estimates and the ground truth values is of the order of $2cm$ for x and y and less than $0.5cm$ for z. From figure 5 we see that the difference between our estimates and the ground truth values is of the order of $2deg$ for $Pitch$ and less than $0.5deg$ for $Roll$ and Yaw.

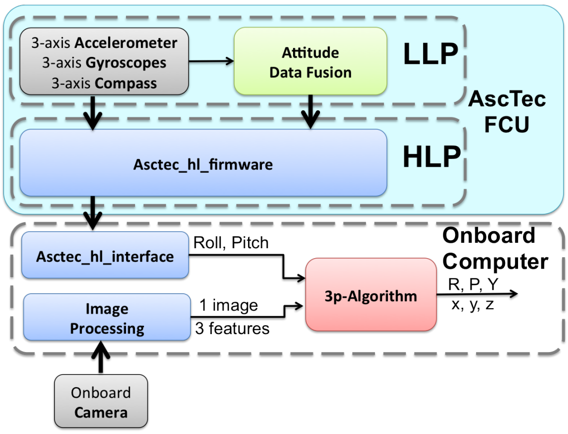


Figure 2. Our Pelican quadcopter: a system overview

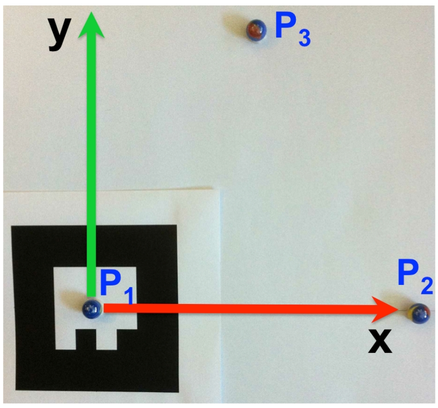


Figure 3. Scenario: The AR Marker and the 3 balls are used only with the aim to get a rough ground truth. The AR Marker provides the camera $6DOF$ pose in a global reference frame according to our conventions.

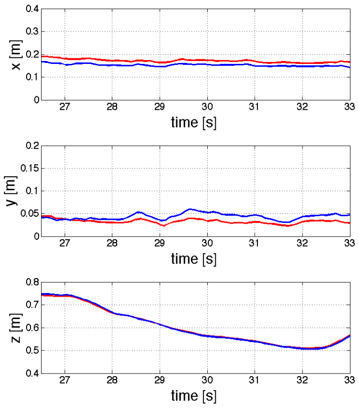


Figure 4. Estimated position, respectively x, y, z. The red lines represent the estimated values with the 3p-Algorithm, the blue ones represent a rough ground truth (from ARToolkit Marker).

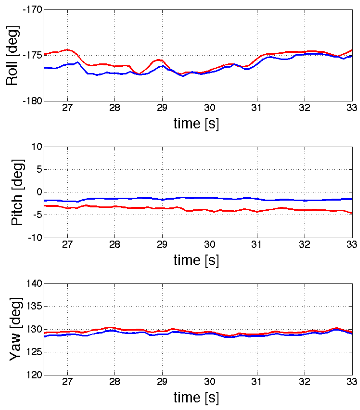


Figure 5. Estimated attitude, respectively $Roll$, $Pitch$, Yaw. The red lines represent the estimated values with the 3p-Algorithm, the blue ones represent a rough ground truth (from ARToolkit Marker).

We believe that the main source of error is due to the distortion of the lens, which is not fully compensated by the calibration. Note that this distortion also affects our ground truth. We plan to test our approach in an environment equipped with a Motion Capture System.

This method is currently under evaluation to be patented.

5.1.2. *Visual recognition for intelligent vehicles*

Participants: Alexandros Makris, Mathias Perrollaz, Christian Laugier.

We developed a generic object class recognition method. The state-of-the-art visual object class recognition systems operate with local descriptors and codebook representation of the objects. Various local features (e.g., gradient maps, edges) are used to create the descriptors. Then kernel based classifiers are commonly employed to classify the detected features in one of several object classes [50] [54]. The recognition of vehicles or pedestrians from sensors mounted on a moving platform is achieved by different approaches using various types of sensors, e.g., stereo camera, laser [61] [52]. The approaches that perform data fusion from various sensors have proven to be the more robust in a variety of road conditions [76].

Our work focuses on the development of an object class recognition system which follows the part based detection approach [65]. The system fuses intensity and depth information in a probabilistic framework. To train the system for a specific object class, a database of annotated with bounding boxes images of the class objects is required. Therefore, extending the system to recognize different object classes is straightforward. We apply our method to the problem of detecting vehicles by means of on-board sensors. Initially, depth information is used to find regions of interest. Additionally, the depth of each local feature is used to weight its contribution to the posterior of the object position in the corresponding scale. The votes are then accumulated in a 3d space-scale space and the possible detections are the local maxima in that space.

The novelty of our approach is the fusion of depth and intensity information to form a probabilistic part-based detector. Using depth information is beneficial for the robustness of the approach, because we avoid including many noisy detections resulting from false matches between features of different scales. The method is tested with stereo video sequences captured in an urban environment. Figure 6 shows some example detections. The proposed method detects cars in various scales, in cases with partial occlusions, and under significant background clutter.

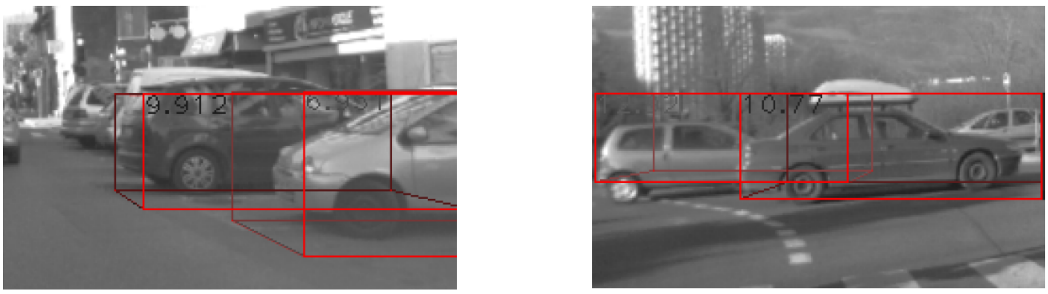


Figure 6. Car detection examples. The new weighting strategy allows to better detect the partially occulded objects.

In 2012, we worked on two particular improvements of the method. First, we modified the weighting strategy in order to increase the detection of partially occulded objects. This approach effectively improves the detection results. Second, we consider replacing the current depth descriptor, which only integrates depth information, with a more advanced depth descriptor (e.g., the NARF descriptor). This work is still in progress, in collaboration with Dimitrios Kanoulas, PhD student in Northeastern University (USA).

In 2012, the full method for objects recognition has been submitted for publication in IEEE Transactions on Intelligent Transportation Systems.

5.1.3. Bayesian Motion Detection in Dynamic Environments

Participants: Qadeer Baig, Jander Perrollaz, Mathias Botelho, Christian Laugier.

5.1.3.1. Introduction

Bayesian Occupancy Filter (BOF) [51] is a grid based perception framework that we use for environment monitoring. In this representation this framework estimates the probability of occupancy as well as velocity of each cell of this grid using sensor data. Output of this framework is used by Fast Clustering Tracking Algorithm (FCTA) [69] to cluster objects and to track them. An important point is that BOF estimates cell velocities without motion information of the ego vehicle, so these are relative velocities. Since no motion information are used, the static objects observed from the moving ego vehicle are also tracked, this results into many false moving objects. Although many of these false positives can be removed by tuning parameters of FCTA, however, this usually is a time consuming task. We note that the number of false can be reduced as well as dependence on FCTA parameters can be relaxed if we can separate the input to BOF into static and dynamic parts. Adding these motion information with cells will allow BOF to calculate velocity information for moving cells only and FCTA will also ignore the static cells while clustering step resulting into faster calculations and better track. In this context we have developed a very fast motion detection technique to separate BOF input into static and dynamic parts. The integration of this module with BOF and FCTA has helped us remove about 78% of the false positives. This technique is summarized next.

5.1.3.2. Fast Motion Detection

In this section we summarize the technique that we have developed to find moving parts of the environment. This motion detection module is situated in the processing chain just before the BOF. The input to this module consists of an occupancy grid generated by the fusion module. And the output of this module is used by both BOF and FCTA modules.

The objective of this module is to separate the input occupancy grid into two parts: cells belonging to static objects and cells belonging to moving objects. The main idea of this separation between static and dynamic parts, consists of keeping a track of how many times a cell is observed as free and how many times it is observed as moving. However to realize this concept we must solve the localization problem. We solve this problem using velocity and rotation information given by MTi-G XSens unit. This allows us to map cells between two input grids OG_{t-1} and OG_t at time $t-1$ and t as shown in figure 7 .

We use two sets of $Free$ and $Occupied$ counter arrays. One set is initialized from new input grid at time t whereas other set keeps updated counts until time $t-1$. Then after above transformation betweens cells of grids OG_{t-1} and OG_t newly initialized set of arrays is updated from arrays at time $t-1$, resulting in incremented counts for overlapping areas between two grids. Finally following decision function is used to separate cells of current input grid OG_t into static and dynamic parts and results are stored in a motion grid.

$$MotionGrid_t[i] = \begin{cases} 1, & OG_t[i] > 0.5 \text{ and} \\ & FreeCount_t[i] > 2 * OccupiedCount_t[i] \\ 0, & \text{otherwise} \end{cases} \tag{34}$$

This technique being simple is quite robust and efficient and does not oblige us to solve the complete SLAM problem. This work is published as [19] and [20].

5.1.3.3. Integration within the BOF framework

We have updated the BOF implementation to take into account the motion detection results. The motion grid is used as an input for updating the BOF. If the input motion grid tells that a cell belongs to a static object, then during prediction and update cycles of BOF the cell's velocity distribution over the velocity range is set to uniform for all discrete velocity values. This essentially means that no velocity information for a given cell

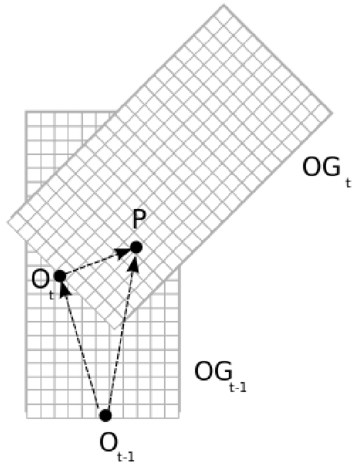


Figure 7. Position of the grid at time instants $t-1$ and t. Vehicle undergoes a motion of $u_t = (\nu_t, \omega_t)$ to move from O_{t-1} to O_t. We need to find the position of point P of grid OG_{t-1} in grid OG_t.

is available and the cell is labeled as static in the current BOF implementation. However, if the cell has been detected as belonging to a moving object, then the velocity distribution prediction and the update cycle are carried out normally. In formal terms this change in the parametric form of dynamic model can be stated as:

$$P(A_i^t | A_i^{t-1}) = \begin{cases} (1-\epsilon)P(A^{t-1}_{A_i^{t-1}}) + \epsilon/\|A_i\| \\ \quad \text{if } MotionGrid_t[i] > 0 \\ \\ 1/\|A_i\| \quad \text{otherwise} \end{cases}$$

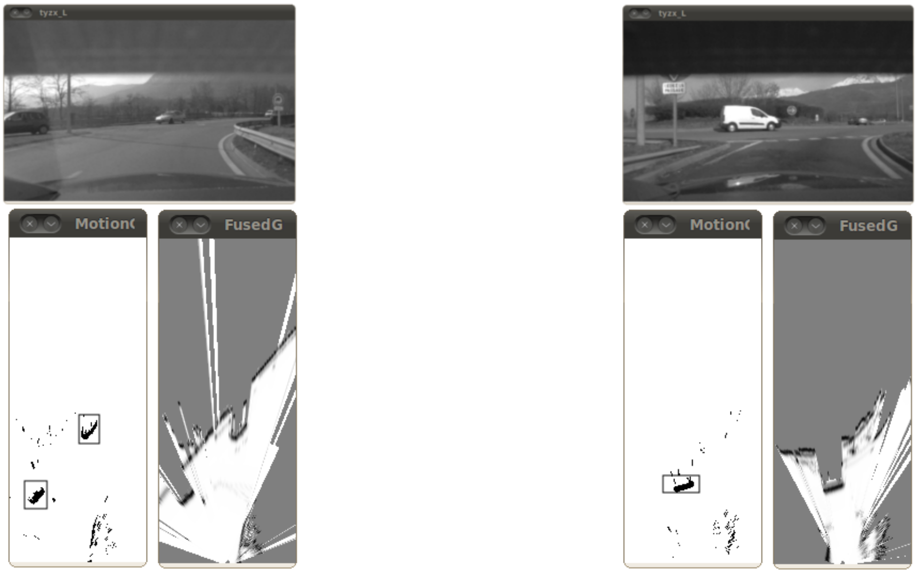
where A_i^t is the set of antecedents of cell i at time t and ϵ is a parameter of BOF, modelling the prediction error probability.

5.1.3.4. Integration with FCTA

We have also updated the FCTA implementation to take into account the motion detection results. The cells which do not possess the velocity information are now ignored during the clustering step. While generally most of the areas belong to static objects and are detected as static by the motion detection module, two main advantages are expected from this strategy: (i) the clustering stage of the algorithm is highly accelerated by the reduction of hypotheses, and (ii) the false moving clusters are ignored because they are not considered for clustering, even with the relaxed FCTA parameters.

5.1.3.5. Results

Some qualitative results of motion detection module are shown in figure 8 , (rectangles around the objects are drawn manually to highlight them). As expected, the moving objects are properly detected. For example, figure 8 (left) shows the motion detection scenario of two cars, and the car moving around a roundabout has been successfully detected in figure 8 (right). Some noise is also visible on the results, mainly due to two causes: first, the uncertainty on the IMU measurements along with the circular motion model may result in some errors in the estimation of the motion; second, the decision function is too rough for taking correct decisions in every situation. The results would benefit from replacing this function by a probabilistic model.



a) b)//

Figure 8. Left: Motion detection results of two cars. Top, scenario, bottom right input fused grid, bottom left resulting motion grid. Right: Motion detection results of a car on a roundabout. Some noise due to sensor uncertainty is also visible

The tracking results of FCTA are highly sensitive to its parameters values. There are less false positives when strict parameters (large thresholds) are used, however, a large number of the true tracks may be missed, resulting in numerous miss detections -note that since the focus of this work is to detect moving objects, we consider in this part that detections belonging to the static environment are false alarms-. The relaxed parameters (small thresholds) provide less miss detections, however, a large number of false tracks are detected. While finding the appropriate set of parameters can be a challenging task, our implementation of the motion detection module with relaxed parameters represents a trade-off.
The following statistics with a dataset duration of about 13 minutes give an insight into the improvements gained with this implementation. When the motion detection module is not used, 22303 tracks are detected. The activation of the motion detection module with all other parameters being equal provides to detect 4796 tracks. This example shows the advantage of the motion detection module because it allows us to remove most of the false tracks while leaving most of the true tracks. Some qualitative FCTA tracking results with and without motion detection module activated (with all other parameters being same) are shown in figures 9 and 10 . Red rectangles are the detected tracks by FCTA in the shown scenario. We clearly see that most of the false positives have been removed.

5.1.3.6. Conclusion

In this section we have presented a fast technique to find moving objects from laser data and its integration with Bayesian Occupancy Filter (BOF) and Fast Clustering-Tracking Algorithm (FCTA). We have seen that after this integration we were able to remove a significant number of false alarms, this has also relaxed the dependence of results on the FCTA parameters.

We plan to change the rather ad hoc decision module that is currently based on occupied and free counter values to a more formal probabilistic function that also takes into account the uncertainty effects on the neighboring cells to accommodate the localization errors. We are also working on extending the tracking module from single motion mode to multiple motion modes.

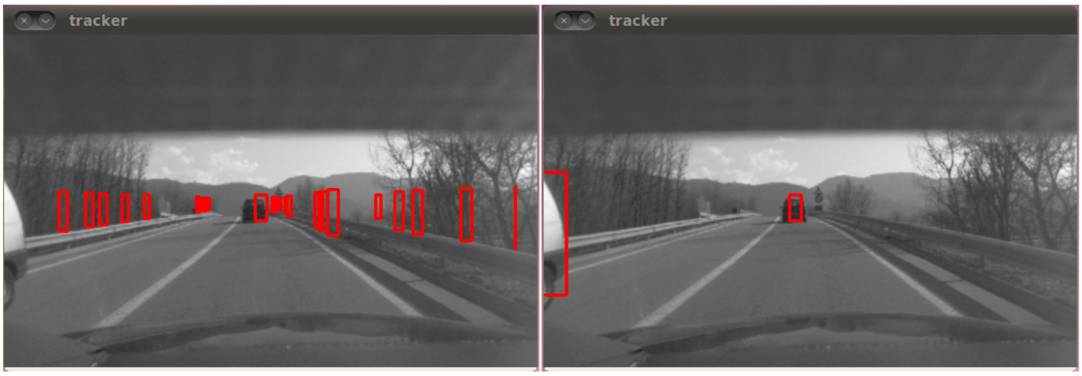


Figure 9. Tracking results of a car. Left, FCTA results without motion detection module activated. Right, same scenario but with motion detection module activated.

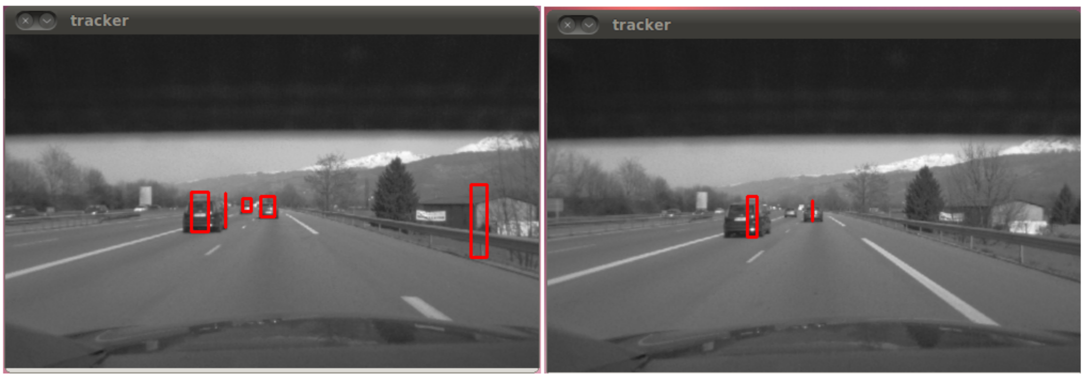


Figure 10. Tracking results of two cars on highway. Left, FCTA results without motion detection module activated. Right, same scenario but with motion detection module activated.

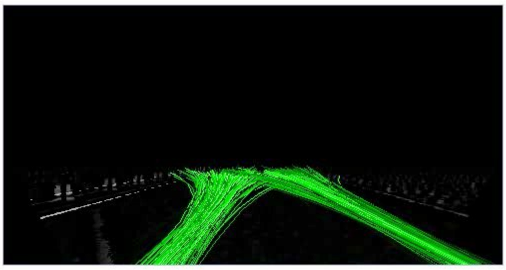
5.1.4. Vision-based Lane Tracker

Participants: Mathias Perrollaz, Amaury Nègre.

For perception in road structured environment the detection of the lane markers and its localization provide an interesting information to predict drivers behaviors and to evaluate collision risks. We currently develop a real time road lane detection and tracking application using camera's image information. The tracking application estimates simultaneously the road plane orientation, the lane curvature and the camera position by using a Monte-Carlo particle filter. With this method, the parameter distribution is represented by a set of particles (see Fig 11 .a) that are sequentially updated using the vehicle dynamic model, evaluated by a ridge extraction (Fig 11 .b) and sampled considering the evaluation result. The average of the particles, displayed on Fig 11 .c) provides a good estimation of the lane state.

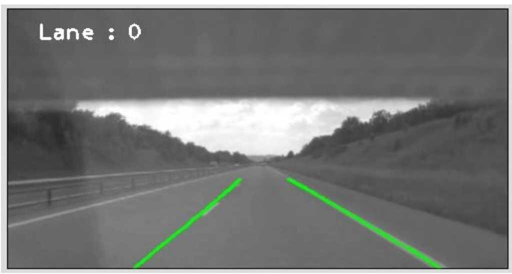
To obtain real-time performance, we implemented the whole process on GPU using the nVidia Cuda toolkit.

The output of this application has been mainly used to predict lane change behaviour 5.2.1 and to risk estimation applications.



(a) Particles cloud

(b) Ridges extraction



(c) Lane state estimation

Figure 11. Visual Particle based lane tracking. (a) The Lane state is estimated by a particles set which is recursively updated, evaluated and resampled. (b) A ridge image is compute to estimate each particle. (c) The average of the particle state provides a good estimation of the lane.

5.1.5. Experimental platform for road perception

Participants: Nicolas Vignard, Mathias Perrollaz, Amaury Nègre.

5.1.5.1. Experimental platform material description

Our experimental platform is a Lexus LS600h car shown in Figure 12 . The vehicle is equipped with a variety of sensors including two IBEO Lux lidars placed toward the edges of the front bumper, a TYZX stereo camera situated behind the windshield, and an Xsens MTi-G inertial sensor with GPS.

Figure 12. Lexus LS600h car equipped with two IBEO Lux lidars, a TYZX stereo camera, and a n Xsens MTi-G inertial sensor with GPS.

The stereo camera baseline is 22 cm, with a field of view of $62°$. Camera resolution is 512x320 pixels with a focal length of 410 pixels. Each lidar provides four layers of up to 200 impacts with a sampling period of 20 ms. The angular range is $100°$, and the angular resolution is $0.5°$. The on-board computer is equipped with 8GB of RAM, an Intel Xeon 3.4 GHz processor and an NVIDIA GeForce GTX 480 for GPU. IMU data contains accelerations, velocity, GPS position and steering angle. The experiments are conducted in various road environements (country roads, downtown and highway), at different time of the day, with various driving situations (light traffic, dense traffic, traffic jams). The datasets are acquired online and are used for testing of our sensor fusion and risk assessment algorithms.

5.1.5.2. Migration from Hugr to ROS middleware

Our platform described in 5.1.5.1 previously used a middleware named Hugr. Middlewares bring an abstraction layer between the sensors drivers and the processing modules. We also used this middleware to share information with modules and applications. Using a middleware facilitates and normalises the communication between modules.

Hugr has been developed by inria for the Cycab project and a team was built to add functionalities and maintain this new middleware. However, now the team has to work on other projects and it is becoming increasingly difficult to allocate resources to maintain this middleware. Given this and some other technical issues [49], we have decided to change our robotic middlware.

We find that many different middleware (AROCAM, RTMaps, ROS,$\cdots$) are being used in the robotic community [53]. Among these, Robotic Operating System (ROS) is increasingly becoming a research standard in robotics. The reason being: an important community, a lot of tools and sharing work and development. The primary goal of ROS is to develop faster robotics applications. However, before moving to ROS we also did an extensive research on the comparison between Hugr and ROS [49], that supported our this decision.

Because of this middleware change, we had to reimplement all the perception process from drivers to applications. In this regard, we have implemented the following drivers:

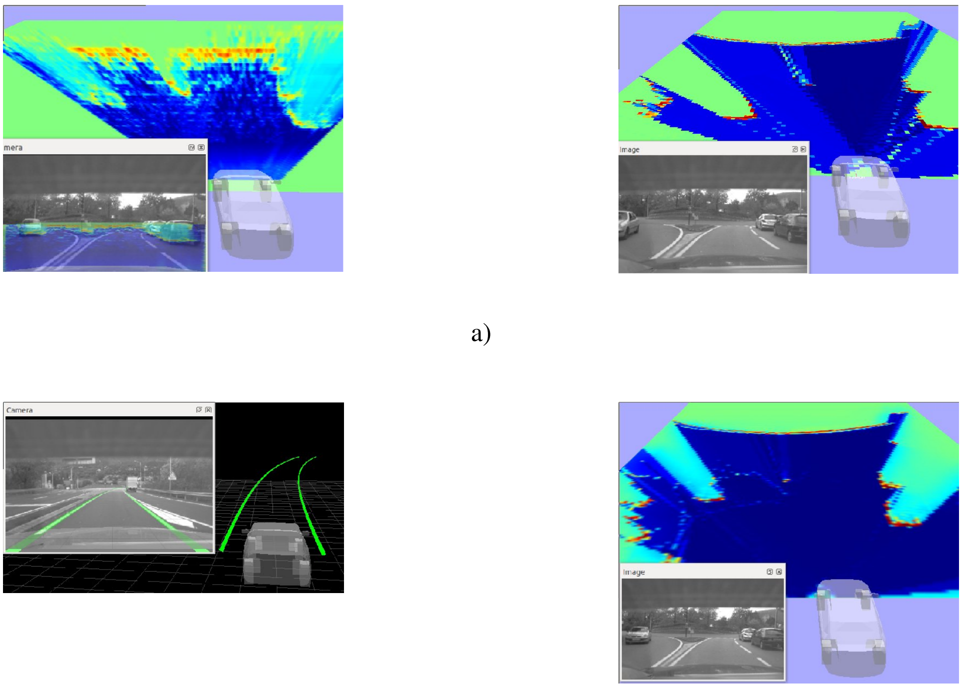
- the IBEO Lux lidar
- the TYZX camera
- the CAN bus
- the Xsens MTi-G (inertial sensor with GPS)

However for the Xsens MTi-G, we found an existing driver that we modified to add the GPS functionality http://www.ros.org/wiki/lse_xsens_mti.

Furthermore, we have also migrated the following modules:

- a module that fuses lidar data into an occupancy grid
- a module that generates occupancy grid from the stereo camera
- the Bayesian Occupancy filter (BOF) module
- the lane tracker

Some result images of occupancy grids and data from the lane tracker after this migration to ROS are shown below 13 . Finally, we have created a public repository at http://gforge.inria.fr that share our developments (both drivers and modules).



a) b)

c) d)

Figure 13. a) occupancy grid from the stereo camera. b) occupancy grid from the lidar. d) lanes detected by lane tracker. e) occupancy grid from the BOF.

5.1.5.3. Disparity space approach for a vision based occupancy grid

Participants: Mathias Perrollaz, Anne Spalanzani, John-David Yoder, Amaury Nègre, Christian Laugier.

To use sensors in the BOF framework, it is essential to develop an associated probabilistic sensor model that takes into consideration the uncertainty over measurements. In 2009, we proposed such a sensor model for stereo-vision [72]. The originality of the approach relied on the decision to work in the disparity space, instead of the classical Cartesian space. In 2010, we improved our sensor model, in order to mimic some features of the sensor models used for range finders. Particularily, we worked on managing visible/occluded areas of the scene [74], and on including the information from the road/obstacle segmentation of the disparity image [73]. Our approach was also designed to allows highly parralel computation of the occupancy grid. A. Nègre implemented the approach on GPU using NVIDIA CUDA to enhance the performance. The complete processing of the stereo data can now be done in 6 ms, while more than 150 ms were necessary with the CPU implementation. The complete approach for occupancy grid computation using stereovision has been publish in 2012, in [13].

5.1.6. *Software and Hardware Integration for Embedded Bayesian Perception*

Participants: Mathias Perrollaz, Christian Laugier, Qadeer Baig, Dizan Vasquez.

The objective of this recently started research work is to re-design in a highly parallel fashion our Bayesian Perception approach for dynamic environments (based on the BOF concept), in order to deeply integrate the software components into new multi-processor hardware boards. The goal is to miniaturize the software/hardware perception system (i.e., to reduce the size, the load, the energy consumption and the cost, while increasing the efficiency of the system).

This work has been started in 2012 in cooperation with CEA-LETI DACLE laboratory. During 2012, we have worked on the definition of the software/hardware architecture and we have started to re-think some components of the lower layer of the BOF software module.

The work plan has been split in two three-year-long phases, respectively leading to address a first level of integration based on mobile technologies, and a second level of integration, based on a more dedicated hardware architecture (and maybe to a SOC).

Two cooperative projects have been prepared and submitted this year for supporting this promising research: the “Permobile” project (FUI), involving industrial companies and user, and the “Perfect” project (IRT-Nano) involving the CEA-LET LIALP lab and ST-Microelectronics. Permobile is focusing on the first integration objectives (3 years) and has been recently submitted. Perfect is focusing onto the second integration objectives (6 years) and the development of integrated open platforms in the domain of transportation (vehicle and infrastructure) and in a second step in the domain of health sector (mobility of elderly and handicapped people, monitoring of elderly people at home…).

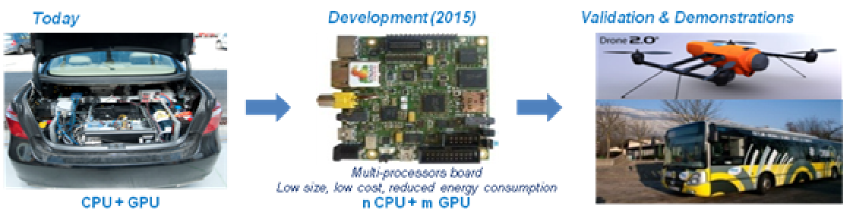


Figure 14. First objective for software/hardware of the BOF: developing and using multiple processor board from mobile technologies. The approach will be validated with real demonstrators.

5.2. Dynamic Change Prediction and Situation Awareness

5.2.1. *Vision-based Lane Change Prediction*

Participants: Puneet Kumar, Mathias Perrollaz, Stephanie Lefevre, Amaury Nègre, Maiwen Gault.

Predicting driver’s behaviors is a key component for future Advanced Driver Assistance Systems (ADAS). In 2012, we have proposed a novel approach for lane change prediction, using only information from a vision sensor embedded into the car. The idea is to predict in advance if our vehicle is about to change lane. Then this information can be used to properly help the driver, for instance by detecting inconsistencies with the turn lights signals.

As an input, the method uses visual data from a camera embedded into the car. A multiple-size ridge filter is used to extract low level features from the image (white markings on black road). Then road lanes are estimated and tracked over time using a particle filter. This process allows parallel computing, and thus works in real time on GPU.

Figure 15. Vision-based tracking of the road markings. From left to right: particles generated by the particle filter, low level features extracted using the ridge filter, and estimated lane.

The road markings are used to estimate the position and heading angle of our car with respect to the lane, as well as the derivatives of these variables. This information is then used as a vector of features for a classifier. The used classifier is a multi-class Support Vector Machine (SVM). The three possible classes are "no lane change" (NL), "right lane change" (CR) and "left lane change" (CL). The classifier has been trained using real data of 180 lane changes on highway, manually annotated. The output of the classification is then converted into a set of probabilities using a generalized Bradley-Terry model.

The classifier provides a very short term classification, which can contain many errors. The longer term integration of the time information is obtained by feeding the classification results into a Bayesian Filter (BF). The posterior output of the filter provides the probability distribution over possible behaviors (NL, CR, CL), hence providing the lane change prediction.

Real-world data from our vehicle is used for the purpose of training and testing. Data from different drivers on different highways were used for the robustness evaluation of the overall approach. The proposed method show promising results, because it is able to predict driver's intention to change lane 1.3 seconds (average) in advance, with maximum prediction horizon of 3.29 seconds. We are now working on a real time implementation of this approach, to demonstrate its use on real situations (e.g., for warning the driver while driving on the highway).

5.2.2. Risk estimation at road intersections for connected vehicle safety applications

Participants: Stéphanie Lefèvre, Christian Laugier.

Intersections are the most complex and dangerous areas of the road network. Statistics show that most road intersection accidents are caused by driver error and that many of them could be avoided through the use of Advanced Driver Assistance Systems. In this respect, vehicular communications are a particularly promising technology. The sharing of information between vehicles over wireless links allows vehicles to perceive their environment beyond the field-of-view of their on-board sensors. Thanks to this enlarged representation of the environment in time and space, situation assessment is improved and dangerous situations can be detected earlier.

A PhD was started on this topic in 2009, in collaboration with Renault. It tackles the problem of risk estimation at road intersections from a new perspective: a Bayesian framework is proposed for reasoning about traffic situations and collision risk at a semantic level instead of at a trajectory level. While classic approaches estimate the risk of a situation by predicting the future trajectories of the vehicles and looking for intersections between them, here dangerous situations are detected by estimating the intentions of drivers and looking for conflicts between them. This novel approach to risk assessment is very relevant in the context of road traffic, as it takes into account the fact that the road network is a highly constrained environment regulated by traffic

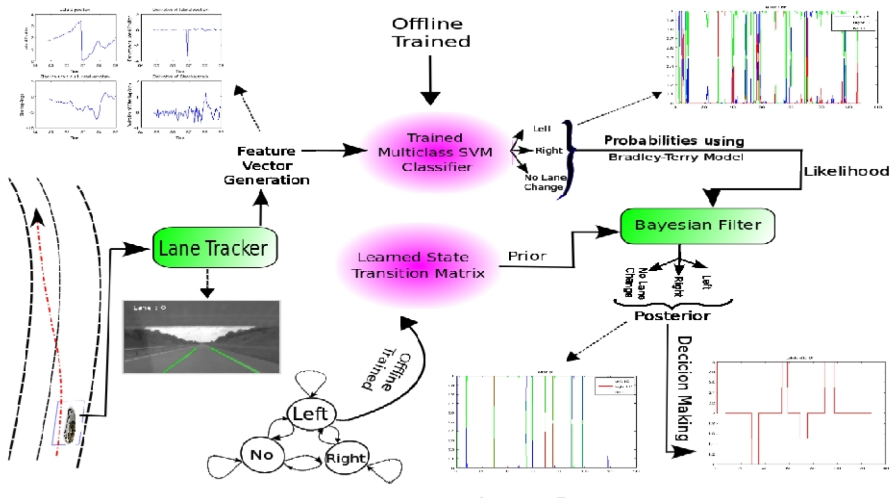


Figure 16. General architecture of the lane change prediction module.

rules. The proposed approach relies on the estimation of drivers' intentions, and the main difficulty lies in the presence of uncertainties in the estimation process: uncertainties inherent to sensor data, and ambiguities when linking vehicle behavior with driver intention. In this work the information about the state of other vehicles is obtained via vehicle-to-vehicle communication, but the proposed framework for reasoning on traffic situations and risk is general and can be applied with other types of sensors, e.g., the ones presented in 5.1.3 .

The focus of the first year (2010) was on estimating a driver's intended maneuver at an intersection (go straight, turn left, etc.) based on the current state of the vehicle (position, orientation, turn signal state) and on contextual information extracted from the digital map. The idea was to use the information on the geometry of the road network and on the connectivity between lanes to build a statistical model of the relationship between a vehicle's state and the driver's intended maneuver. The proposed solution is based on a Bayesian Network and on geometric functions which automatically extract the characteristics of the intersection from a digital map. This approach was designed and implemented during a 3-month internship in the Stanford Artificial Intelligence Laboratory, in collaboration with Sebastian Thrun's Driving Group.

During the second year (2011) we augmented the Bayesian Network with a filtering process so that new measurements could be recursively used to estimate the driver's intentions. The new version of the motion model explicitly models the influence of traffic rules on the behavior of a vehicle. While state-of-the-art approaches usually assume independence between vehicles, the proposed motion model takes into account the mutual influences between the maneuvers performed by the vehicles in the scene. These improvements were carried out by introducing two new variables in the Bayesian Network. The ***"Intention to stop"*** corresponds to the driver's intention to come to a halt at the intersection. The ***"Expectation to stop"*** corresponds to whether or not the traffic rules expect the driver to come to a halt at the intersection. The former is assumed to be dependent on the previous intention of the driver and on the current expectation. The latter is assumed to be dependent on the rules applying at the intersection and on the previous situational context, i.e., the state of the other vehicles in the scene. With this model it is possible to infer what a driver intends to do and what a driver is expected to do from the successive measurements of the pose, speed, and turn signals of the vehicles in the scene. Risk can then be computed based on the probability that intention and expectation do not match.

The focus of this year (2012) was on the evaluation of the performance of the algorithm. The proposed approach was validated in field trials using passenger vehicles equipped with vehicle-to-vehicle wireless communication modems, and in simulation. Our simulations assumed ideal perception and communication, and considered typical accident scenarios at a two-way-stop cross intersection. The tested maneuvers included crossing maneuvers, merging maneuvers, and left turn across path maneuvers (see Figure 17). A total of 240 instances of these scenarios were simulated, with both priority violations and stop violations as accident causes. The same number of instances were simulated for non-dangerous situations, by enforcing a 3 seconds safety distance between the vehicles at all times. An analysis of the collision prediction horizon led to the following conclusions:
1. There were no false alarms in non-dangerous situations, and no missed detection in the dangerous scenarios.
2. For merging and crossing maneuvers, the proposed algorithm was able to predict collisions at least 1.5 s before they occurred.
3. For left turn across path maneuvers, the proposed algorithm was able to predict collisions at least 0.6 s before they occurred.
4. Accidents caused by stop violations were detected on average 1 s earlier than the ones caused by priority violations.
Different accident avoidance strategies were tested: warning the driver of the vehicle with right-of-way, warning the driver of the other vehicle, applying autonomous braking on the vehicle with right-of-way, and applying autonomous braking on the other vehicle. It was found that the ability of each strategy to avoid an accident varies a lot with the situation. For example, the "autonomous braking on the vehicle with right-of-way" can avoid the accident in 91% of cases for stop violations, but only in 34% of cases for priority violations. "Warning the driver of the vehicle with right-of-way" can avoid the accident in 1% of cases for priority violations, while for the same scenarios "autonomous braking on the other vehicle" can avoid the accident in 99% of cases. These results were published at the conference IEEE IROS'12 [22], and as an Inria Research Report [41]. Field trials were conducted using two vehicles equipped with off-the-shelf vehicle-to-vehicle wireless communication modems. Six different drivers took part in the experiments to recreate realistic dangerous and non-dangerous situations at a T-shaped give-way intersection (see Figure 17). The risk estimation algorithm was run online in one of the vehicles, and triggered a warning for the driver when it detected a dangerous situation (see Figure 18). In the 120 tests, there were no false alarms and no missed detections. The warning was always triggered early enough that accidents were avoided by performing an emergency braking. The field trials proved that the proposed approach can operate with success in real-life situations and trigger warnings in real time. They also showed the robustness of the algorithm, since the experiments were carried out with several drivers, a positioning system with a precision of 2 meters (standard deviation) and challenging wireless communication conditions. These results were published at the conference IEEE IV'12 [23], where the paper received the Best PhD Student Paper award.

The PhD was successfully defended in October 2012 [9]. A patent application was filed with Renault in October 2012 [45]. This work will be continued within the Inria@SiliconValley program, in collaboration with the University of Berkeley, California. Ms Lefevre will conduct further research on this topic as a post-doctoral researcher at Berkeley starting January 2013.

5.2.3. Guidance for Uncertain shooting domain

Participant: Emmanuel Mazer.

This study is made in collabortion with MBDA (Monsieur Le Menec) and Probayes (Monsieur Laurent Saroul) under the ITP framework financed by the british MOD and the french DGA

Contex This project relates to the use of lock after launch missiles, both long range anti aircraft missiles such as Meteor, or air to ground strike weapons employing for example IIR or Semi Active Laser (SAL) guidance. In both cases, a target is ultimately recognized and tracked by means of a seeker which detects a characteristic signal above the noise. This could be the target reflections of a radar beam, or the spot from a designating laser.

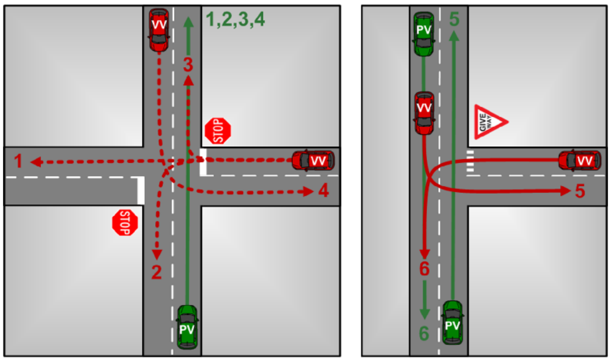


Figure 17. Scenarios tested in simulation (left) and during field trials (right).

Figure 18. Online execution of the algorithm during the field trials: warning the driver of an upcoming collision with a vehicle on the left.

However, a missile is often launched at a target range which is greater than its seeker acquisition range, although within the kinematics No Escape Zone (NEZ). It is provided with targeting geometry before launch, and maybe (via a data link) during the first part of the trajectory. However, it must fly for some period in inertial mode, and during this time the target may manoeuvre. Also, errors build up due to the imperfections in the inertial navigation system. This means that the target bearing becomes increasingly uncertain whilst the range reduces. It may be necessary to scan the seeker to acquire the target. If the scan is not matched to the possible manoeuvres, the target may escape detection. But if the scan is large, the acquisition range will be reduced, because of the reduction in search time per solid angle. As the target is acquired later, the missile's terminal manoeuvre will be more severe, and as a result the range assumed for the original kinematics NEZ may have been too optimistic. Equivalently, it is possible to be too pessimistic about the target uncertainty, hence to scan too much, and acquire the target so late that there is no longer the manoeuvre capability to reach it. Present Weapon systems optimize the probability of successful interception assuming either Gaussian uncertainties, or worst case uncertainties.

Objectives and achievments of the GUS-D system

These considerations lead to the concept of a stochastic approach for computing a probabilistic, adaptive NEZ. Probabilistic NEZ depend on the uncertain target behaviour. The uncertainties we propose to deal with are also related to the missile Inertial Navigation System (INS) precision, to sensor errors and to misalignments. Moreover, the uplink management, i.e., when to evade and breakdown the link between the launching platform and the in-flight missile plays a major role on the target localization accuracy and by the way to the size of the NEZ. Finally, there is uncertainty in the target radar cross section, which has a big effect on the seeker acquisition range. The purpose of this study is better tactical advice to the pilot about launching decision and how long maintain the uplink, and where appropriate, better matching of seeker scan strategies to target behaviours. These decisions have impacts on the probability of combat success; i.e., not only to hit the target but also on the probability to survive, as the opponent aircraft or ground threat may launch similar weapons.

The project focuses predominantly on Air to Air systems. The Meteor scan strategy has been studied deeply and is no longer critical for the engagement of fighter jets, but an objective of the study is to extend the strategy to the engagement of targets of much lower radar cross section, where the acquisition range is significantly shorter. Nevertheless all the issues apply also to Air to Ground weapon systems.

The GUS-D system is limited to one to one engagement scenario:

- one aircraft and its missile
- opponent aircraft and its missile

The main functionality of the GUS-D system is then to provide to the user a probability of successful target interception given the current engagement conditions, and the uncertainties on the target properties and behaviours.

5.3. Human Centered Navigation in the physical world

5.3.1. Goal oriented risk based navigation in dynamic uncertain environment

Participants: Anne Spalanzani, Jorge Rios-Martinez, Arturo Escobedo-Cabello, Procopio Silveira-Stein, Alejandro Dizan Vasquez Govea, Christian Laugier.

Navigation in large dynamic spaces has been adressed often using deterministic representations, fast updating and reactive avoidance strategies. However, probabilistic representations are much more informative and their use in mapping and prediction methods improves the quality of obtained results. Since 2008 we have proposed a new concept to integrate a probabilistic collision risk function linking planning and navigation methods with the perception and the prediction of the dynamic environments [57]. Moving obstacles are supposed to move along typical motion patterns represented by Gaussian Processes or Growing HMM. The likelihood of the obstacles' future trajectory and the probability of occupation are used to compute the risk of collision. The proposed planning algorithm, call RiskRRT (see Figure20 for an illustration), is a sampling-based partial planner guided by the risk of collision. Results concerning this work were published in [58] [59] [60]. In

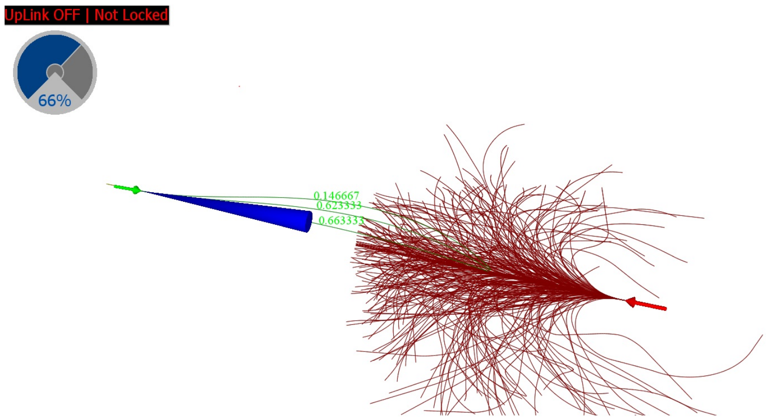


Figure 19. This Figure shows (up left corner) the "probability meter" which indicates the probability to intercept the target if the uplink is to be shutoff now. The 3D representation of the scene is displayed. The red filaments are future possible trajectories obtained with a Markov Process (MBD-UK). The blue cone modelizes the detection cone of the seeker. The green numbers indicates the probabilities to lock the target with the corresponding trajectory.

2012, We continue to work on developing probabilistic models and algorithms to analyze and learn human motion patterns from sensor data (e.g., tracker output) in order to perform inference, such as predicting the future state of people or classifying their activities. Our work has been published in the Handbook of Intelligent Vehicles [40]. We obtained some preliminary results on our robotic wheelchair combining RiskRRT with some social conventions described in section 5.3.2 . This approach and experimental results have been published at ISER 2012 [32].

This algorithms is used in the work presented in the next three sections, work conducted under the large scale initative project PAL.

5.3.2. Socially-aware navigation

Participants: Jorge Rios-Martinez, Anne Spalanzani, Alessandro Renzaglia, Agostino Martinelli, Christian Laugier.

Our proposal to endow robots with the ability of socially-aware navigation is the Social Filter, which implements constraints inspired by social conventions in order to evaluate the risk of disturbance represented by a navigation decision.
The Social Filter receives from the perception system a list of tracked humans and a list of interesting objects in the environment. The interesting objects are designated manually according to their importance in a particular context, for example, an information screen in a bus station. After the process of such data, the Social Filter is able to output the risk of disturbance relative to people and interesting objects, on request of the planner and the decisional system. Thus, the original navigation solutions are "filtered" according to the social conventions taken into account. Notice that the concept of social filter is built as a higher layer above the original safety strategy, the planner and the decisional system are responsible to include the new constraints.
The on-board Kinect attached to our robotic platform was used to track people and to detect interactions. The Kinect sensor permits to get the position and orientation of the torso for each identified human. That information is passed to the Social Filter. Result images can be seen in Figure 21 .

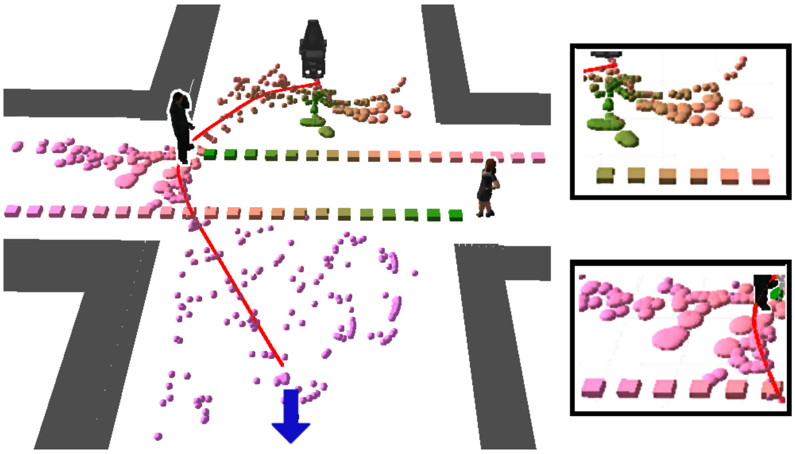
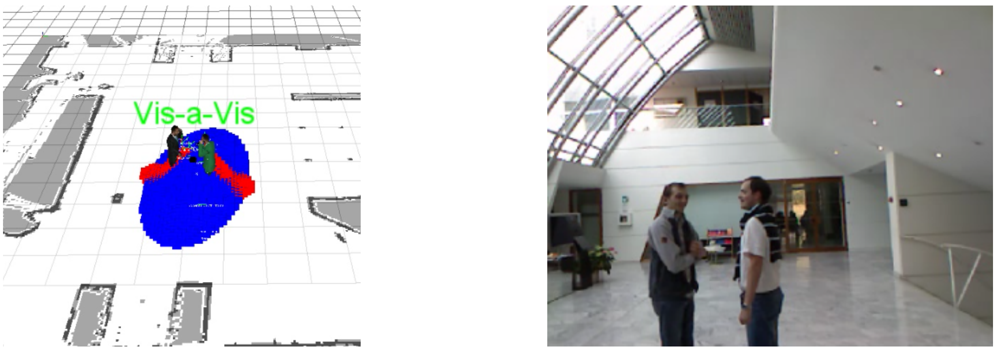


Figure 20. Predictive navigation example. RiskRRT selected a plan (red line) to the goal (blue arrow). The chosen path leads the robot to pass by the back of the first person, and then reduces the speed to let the second person to pass as well. With this strategy, the robot minimizes the risk of collision and the discomfort caused for the two pedestrians. Once second person has passed, the algorithm choses a straighter path to the goal. Frames at the right of the figure show that estimated risk is bigger at future positions of the wheelchair (circles) which are close to predicted positions of pedestrians (squares).



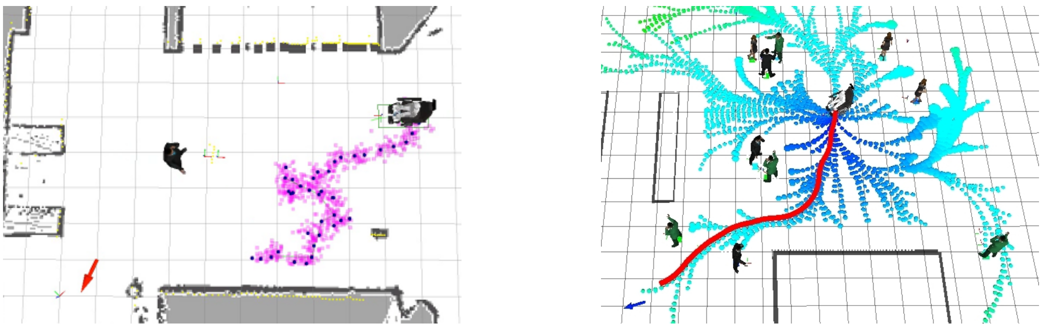
(a) (b)

Figure 21. Interaction detected with Social Filter from Kinect input for a pair of humans. Torso direction is used to estimate the main focus of interest.

In the context of socially-aware robot navigation in dynamic environments, as part of Jorge Rios-Martinez PhD thesis (to be defended in january 2013), two techniques have been proposed: one considering optimization-based navigation presented in [26] and the other a Risk-based navigation approach, previously presented in [75].

The **optimization-based navigation strategy**, done in collaboration with A. Renzaglia, is based on the Cognitive-based Adaptive Optimization (CAO) approach applied to robots [10]. We formulate the problem of socially-aware robot navigation as an optimization problem where the objective function includes, in addition to the distance to goal, information about comfort of present humans. CAO is able to efficiently handle optimization problems for which an analytical form of the function to be optimized is unknown, but the function is available for measurements at each iteration. A model of social space, contained in the Social Filter module, was integrated in order to work as a "virtual" sensor providing comfort measures. Figure22 a) shows an image of the method implementation on ROS [2] framework.

Social Filter models of social conventions were combined with RiskRRT [56] by including the knowledge of human management of space (Personal Space, interaction space, activity Space). The particular considered interaction was the conversation between pedestrians which was missed in the most part of related works. The approach presented shows a way to take into account social conventions in navigation strategies providing the robot with the ability to respect the social spaces in its environment when moving safely towards a given goal. Due to the inclusion of our social models, the risk calculated for every partial path produced by RiskRRT algorithm is given by the risk of collision along the path and the risk of disturbance to human spaces.



(a) (b)

Figure 22. Results of socially-aware navigation approaches. In a) the optimization-based navigation solution avoids a region were the discomfort for the human would be higher. In b) the Risk-Based navigation technique explores the space and decides to follow a path avoiding social spaces minimizing the risk of disturbance. The goal in each case is signaled by an arrow.

One last work was presented in [25], where the socially-aware navigation based on risk was integrated with a model of human intention estimation (presented in section 5.3.4 . Results exhibited emerging behavior showing a robotic wheelchair interpreting facial gesture commands, estimating the intended goal and autonomously taking the user to his/her desired goal, respecting social conventions during its navigation.

5.3.3. Navigation Taking Advantage of Moving Agents

Participants: Procopio Silveira-Stein, Anne Spalanzani, Christian Laugier.

[2] http://www.ros.org

Following a leader in populated environments is a form of taking advantage of the motion of the others. A human can detect cues from other humans and smartly decide in which side to pass. Humans can also easily predict the motion of the others, changing his/her path to accommodate for conflictive situations, for example. Imitating the motion of a human can also improve the social acceptance of robots and so on.

The best leader is the one whose goal is close to the robot's one. To implement that, the Growing Hidden Markov Model (GHMM) technique is used [79]. This technique provides at the same time a capability to learn and modeling typical paths, as well as learning and predicting goals associated to paths, making it ideal for the proposed approach of leader election.

Once a leader is chosen, the robot starts to track his/her path and follow it, using the RiskRRT algorithm presented in section 5.3.1 . This algorithm takes into account the risk of collision with other agents, guaranteeing that the robot can avoid collisions even if its leader is lost or occluded.

Some results can be seen in the following experiments, where real human data was used together with a robot simulator.

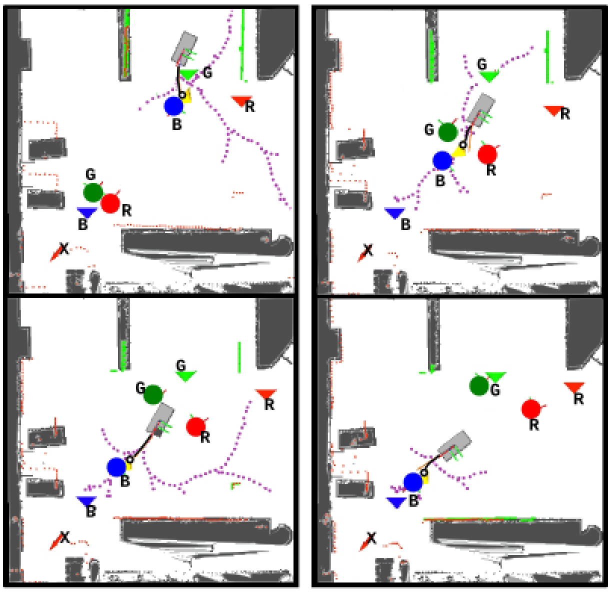


Figure 23. Robot navigation following a leader seamlessly avoid other incoming agents.

In Figure 23 , the experiment demonstrates one of the advantages of following a leader to improve the robot's navigation capabilities. The direct path to the robot's goal is obstructed by two incoming humans. Normally an algorithm suited for dynamic environment would create a detour as future humans' position would conflict with the robot straight trajectory. However, as the robot is following a leader, it does not reason about the other agent's future position. Therefore, the leader knows that people will give room for he/she to pass, and the robot profits from it.

Next step will be to use this technics will navigating in a crowd, task that a common planning strategy could hardly do.

5.3.4. Autonomous Wheelchair for Elders Assistance

Participants: Arturo Escobedo-Cabello, Gregoire Vignon, Anne Spalanzani, Christian Laugier.

The aging of world's population is bringing the need to provide robotic platforms capable to assist elder people to move [77]. It is necessary that such transportation is reliable, safe and comfortable. People with motor disabilities and elders are expected to benefit from new developments in the field of autonomous navigation robotics. Autonomously driven wheelchairs are a real need for those patients who lack the strength or skills to drive a normal electric wheelchair. The services provided by this kind of robots can also be used to provide a service of comfort, assisting the user to perform difficult tasks as traversing a door, driving in a narrow corridor etc. Simple improvements of the classical powered wheelchair can often diminish several difficulties while driving. This idea of comfort has emerged as a design goal in autonomous navigation systems, designers are becoming more aware of the importance of the user when scheming solution algorithms. This is particularly important when designing services or devices intended to assist people with some disability.

In order for the robot to have a correct understanding of the intention of the user (when moving around) it is necessary to create a model of the user that takes into account his habits, type of disability and environmental information. The ongoing research project is centered in the understanding of the intentions of the user while driving an autonomous wheelchair, so that we can use this information to make this task easier.

In 2011 a robotic wheelchair was set up as experimental platform. Some basic functions were included as the mapping of the environment using a Rao-Blackwellized Particle Filter [62], localization using an Adaptive Monte Carlo Localization approach (AMCL) [78], global planning using an A* algorithm [63] and local reactive planning using the Dynamic Window Algorithm [55]. Alongside some work was done with the kinect sensor in order to detect and track people. This behaviour was aimed to bring assistance not only to the user but also to the caregiver by allowing him to move more freely. The software implementation of the related approaches was done on the basis of the ROS middleware.

During 2012 the work was centered in the improvement of the usability of the system around three main axes:

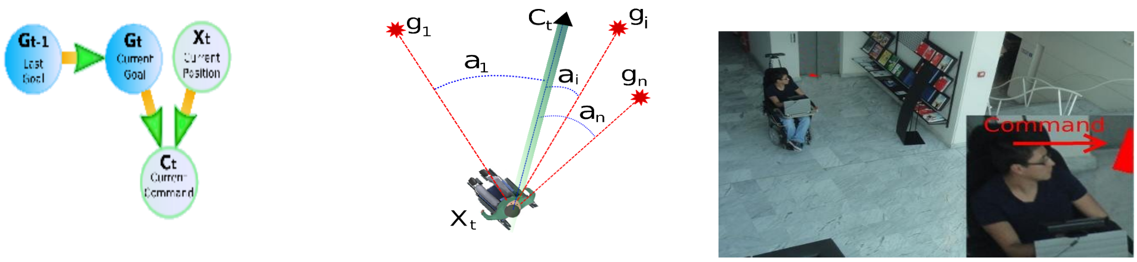
- User intention estimation: A review of the state of the art in user's intention estimation algorithms was made and a new model to infer the intentions of the user in a known environment was presented [46],[47]. The algorithm models the intention of the user as 2D topological goals in the environment. Those places are selected according to how frequently they are visited by the user (user habits). The system was designed so that the user can give orders to the wheelchair by using any type of interface, as long as he can show the direction of the intended movement (joystick, head tracking, brain control, etc). As shown ni figure24 , the chosen approach uses a Bayesian model to model and infer the intentions. The main contribution of this work is to model the intention of the user as topological goals instead of normal trajectory-based methods, therefore the model is simpler to deal with. Current research is being done to understand which information is important to take into account in order to do better estimations of the user's intention. In particular, the movements of the head are considered by the proposed inference method.

  The navigation is performed using the human-aware planning algorithm developed by the team which integrates a notion of social conventions and avoidance of dynamic obstacles to prevent uncomfortable situations when the wheelchair is navigating among humans (see section 5.3.2 for details)

- Interfaces: People with motor disabilities and elders often have problems using joysticks and other standard control devices. Under this consideration our experimental platform was equipped with different types of user-interfaces to provide a multimodal functionality as described in [47]. A face pose interface allows to control the wheelchair's motion by changing the face direction, while voice recognition interface is used to guarantee an adequate control of the wheelchair for those commands that otherwise would be difficult to give by only using the face (Stop, start, etc). The use of a touch screen control is also possible.
- Multimodal control: The wheelchair can be controlled in semi-autonomous mode employing the user's intention estimation module, described later, or in manual mode in which the user is in charge of driving by him self.

  In manual mode the user controls the wheelchair's angular speed moving her head while the linear speed is controlled with vocal commands (faster, slower, break, etc).

In semi-autonomous mode the user shows the direction to his/her desired destination facing towards it. Whenever a new command is read from the face pose estimation system. The user's intention module computes the goal with the highest posterior probability. The navigation module receives the map of the environment, the list of humans present in the scene and the currently estimated goal to compute the necessary trajectory to the goal.



(a) (b) (c)

*Figure 24. **left**:User's intention model. The Bayesian network used to estimate the current user's intended goalG_t, The current position X_t and the user command C_t are used as evidence. G_t is dependent on the value of the last estimation G_{t-1}. **center**: Experimental evaluation of the user's intention module. The probability value for a given command C_t (big arrow) is proportional to the angle a_i formed respect to each goal g_i in the environment. **right:** The user is looking to the left (in the direction of his desired goal). Once that the user's intention estimation system computes the goal with the highest probability, the autonomous navigation module plans the path and controls the movement of the wheelchair to take the user to the destination.*

5.3.5. Multi-Robot Distributed Control under Environmental Constraints

Participants: Agostino Martinelli, Alessandro Renzaglia.

This research is the follow-up of a study begun three years ago in the framework of the European project sFly. The problem addressed is the deployment of a team of flying robots to perform surveillance coverage mission over an *unknown* terrain of complex and non-convex morphology. In such a mission, the robots attempt to maximize the part of the terrain that is visible while keeping the distance between each point in the terrain and the closest team member as small as possible. A trade-off between these two objectives should be fulfilled given the physical constraints and limitations imposed at the particular application. As the terrain's morphology is unknown and it can be quite complex and non-convex, standard algorithms are not applicable to the particular problem treated in this paper. To overcome this, a new approach based on the Cognitive-based Adaptive Optimization (CAO) algorithm is proposed and evaluated. A fundamental property of this approach is that it shares the same convergence characteristics as those of constrained gradient-descent algorithms (which require perfect knowledge of the terrain's morphology and optimize surveillance coverage subject to the constraints the team has to satisfy). Rigorous mathematical arguments and extensive simulations establish that the proposed approach provides a scalable and efficient methodology that incorporates any particular physical constraints and limitations able to navigate the robots to an arrangement that (locally) optimizes surveillance coverage.

Special focus has been devoted to adapt this general approach in order to deal with real scenarios. Specifically, this has been carried out by working in collaboration with the ETHZ (Zurich). To this regard, the approach has been adopted in the framework of the final demo of the sFly project. The demo simulates a search and rescue operation in an outdoor GPS-denied disaster scenario. No laser, no GPS, and Vicon or other external cameras are used for navigation and mapping, but just onboard cameras and IMUs. All the processing runs onboard, on a Core2Duo processing unit. The mission consists of first collecting images for creating a common global map of the working area with 3 helicopters, then engaging positions for an optimal surveillance coverage of the area, and finally detecting the transmitter positions.

The results of this research have been published in two journals, [14], [15], and on the thesis of A. Renzaglia, [10].

5.4. Bayesian Modelling of Sensorimotor Systems and Behaviors

Results described in this section were done in collaboration with the LPPA collège de France.

5.4.1. Bayesian based decision making in multi-player video games

Participants: Gabriel Synnaeve, Pierre Bessière.

The problem addressed in this work is the autonomous replacement of a human player. It is the continuation of last year's work on the same topic as well as a follow-up of previous E-Motion Ph.D Ronan Le Hy [64]. This year, we focused on real-time strategy (RTS) games, in which the players have to build an economy, advance technology, produce and control an army to kill the opponents. From a research point of view, multi-player games are interesting because they stand for a good in-between of the real world and simulations. The world is finite and simulated (no sensors problems) but we didn't wrote the simulation and the other players are humans (or advanced robots in the case of AI competitions).

This year's research work focused on tactical prediction and decision-making as well as armies composition adaptation. For the tactical model, the idea is to have (most probably biased) lower-level heuristics from units observations, which produce information exploitable at the tactical level, and take advantage of strategic inference too. We abstract space into automatically extracted choke points and regions of StarCraft maps from a pruned Voronoi diagram (using [71]). We then assign different scores to each of these regions and learn the influence of these scores on different attack types and locations. To do that, we set up a huge data-set of professional player's games, whose game state was extracted [29]. This work was accepted for publication at Computational Intelligence in Games (IEEE CIG) 2012 in Grenada [30] and was presented at the Computer Games Workshop of the European Conference of Artificial Intelligence (ECAI) 2012 [28].

Another focus of work this year was on army composition adaptation. RTS games unit types combinations in armies can be seen as complex (soft max) rock-paper-scissors games. Our analysis boiled to down army compositions encoded as clusters (we used a Gaussian Mixtures Model) of "classic" combinations (because of economy and technology constraints during the game). This work was published at the AI in Adversarial Real-Time Games workshop of AAAI AIIDE 2012 [29].

On top of the research/evaluation implementation, we also implemented it in our StarCraft: Broodwar's bot implementation BroodwarBotQ. With this bot, we took part in AIIDE and CIG conferences AI tournaments placing respectively 4th (out of 10) and 6th (out of 10). Gabriel Synnaeve defended his thesis on October 24th 2012.

5.4.2. Bayesian modelling to implement and compare different theories of speech communication

Participants: Raphael Laurent, Pierre Bessière, Julien Diard, Jean-Luc Schwartz.

A central issue in speech science concerns the nature of representations and processes involved in communication. The search for phoneme or syllable specific invariants led to three major sets of approaches: motor, auditory and perceptuo-motor theories. They have been widely argued for and against, but the theoretical debate appears to be stagnating. It is our belief that computationnal models designed within a rigorous mathematical framework may allow to put forward new arguments to support either theory, and new ideas for experiments to be carried out on human subjects.

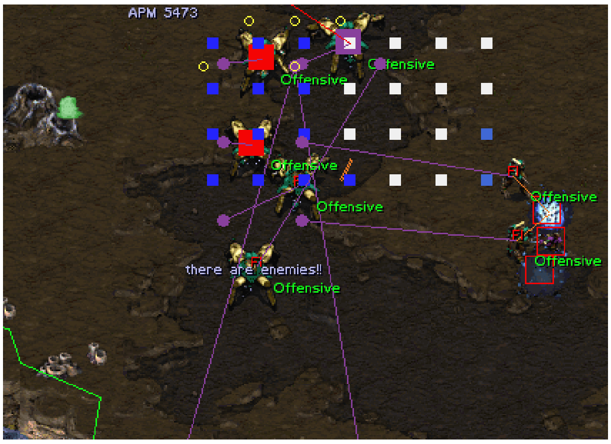


Figure 25. Units movement debugging output during a StarCraft game of the BroodwarBotQ bot. Considering the unit in the upper middle of the picture, white squares represent the highest probabilities of directions, while the darker the blue, the lower the probability to go there. The unit controller searches both to minimizes collisions and stay in range of the enemy targets (bottom right).

We have designed an integrative Bayesian model which allows to study auditory, motor and perceptuo-motor aspects of speech production and perception. In 2011, this model was used to work on purely theoretical simulations where we studied with diverse paradigms the decrease in the performances predicted by the different theories due to communication noise. This work led to the proof of an indistinguishability theorem : given some hypotheses on the learning process, purely motor and purely auditory models have identical answers to perception tasks. Thanks to VLAM, a vocal tract simulation tool which allows to map articulatory parameters to acoustic signals, we tested our model on vowel perception tasks. The results of both these studies are detailed in [70].

In 2012, we worked on a much more complex version of the model, which mas made able to deal with plosive syllable production and perception. A first version of this model was tested on perception tasks on evaluation corpora with more and more variability compared to the learning corpus. This showed a really high robustness of the purely motor model, which contained more information that it is the case in practise, due to unrealistic learning methods. That's why the work was then focused on more realistic learning algorithms, where speech motor gestures are unsupervisedly learned through imitation, by generating motor gestures trying to reach auditory targets, and memorising the acoustics corresponding to these motor commands.

5.4.3. Bayesian programming : book and software

Participants: Emmanuel Mazer, Pierre Bessière.

5.4.3.1. A need for a new computing paradigm

Bayesian probability theory is a mathematical alternative to logic.

However, we want working solutions to incomplete and uncertain problems. Consequently, we require an alternative computing framework based on Bayesian probabilities.

To create such a complete computing Bayesian framework, we require a new *modeling methodology* to build probabilistic models, we require new *inference algorithms* to automate probabilistic calculus, we require new *programming languages* to implement these models on computers, and finally, we will eventually require new *hardware* to run these Bayesian programs efficiently.

Our ultimate goal is a Bayesian Computer. The purpose of this book is to describe a formalism and a computer langage as first steps in this direction.

5.4.3.2. Outline of the book

Its purpose is to introduce the fundamental concepts of Bayesian Programming, to present the novelty and interest of the approach, and to initiate the reader to the Bayesian modeling. Numerous simple examples of applications are presented in different fields.

It is divided in three parts, chapters *2 Basic-Concepts* to *6: Bayesian-Program* which presents the principles of Bayesian Programming, chapters *7 : Information-Fusion* to *11 : Bayesian-Programming-Iteration* which offer a cook book for the good practice of probabilistic modeling and *12 : Bayesian Programming Formalism* to *16 Frequently Asked Question* which revisit the Bayesian inference and learning problems with the help of the presented formalism.

A fist version of the book will be send to the reviewer selected by the editor before the end of 2012

5.4.3.3. Distributed Software

One way to read this book and learn bayesian programming is to run and modify the programs given as example. A Python package "pypl" based on Probt bindings is made available with book.

The source code of the examples as well as the Python package can be downloaded free of charge.

Many examples in the book are given with parts of real corrresponding programs which could be run using the distributed package. They are given under the following format

The figure 26 has been generated using the program "chapter7/invpgm.py". The following instruction allows to get to most probable value for the heading H given the readings.

```
PH=PHkB0B1.instantiate(sensor_reading_values)
best=PH.compile().best()
```

Bayesian programs are also used to generate the illustration of the book such as this one which illustrating the navigation based on sensor fusion.

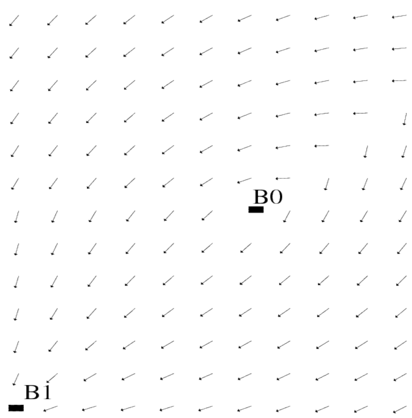


Figure 26. The vector field corresponding to $\max_h P\left(H = h | b_0 \wedge b_1 \wedge \pi\right)$

EXMO Project-Team

6. New Results

6.1. Ontology matching and alignments

We pursue our work on ontology matching and alignment support with contributions to evaluation and alignment semantics.

6.1.1. Evaluation

Evaluation of ontology matching algorithms requires to confront them with test ontologies and to compare the results. Since 2004, we run the Ontology Alignment Evaluation Initiative (OAEI) which organises evaluation campaigns for assessing the degree of achievement of actual ontology matching algorithms [2].

This year, we ran two evaluation campaigns named 2011.5 and 2012. This was justified by the will to complete full evaluations using the support of the SEALS platform. Hence, the main activities carried out in 2012 were related to the automation of the evaluation. This involved providing participants with a better way to bundle their tools so that they can be evaluated both offline and within the SEALS platform. It also required to support more organisers to provide test case within the plaform.

This work has been used in the OAEI 2012 evaluation campaign. OAEI 2012 offered 9 different test sets (7 of which under the SEALS platform). This issue brought the following results:

- More participants than ever (21);
- All ontology matchers running on the SEALS platform (18);
- Increased performances in terms of precision and recall;
- Matchers are now very scalable and can deal with the largest available ontologies (9 systems able to deal with the very large medical ontology SnoMed);

We have also introduced as a data set, the benchmark for multilingual ontology matching developed last year [6]. It has pushed matcher developers to address multilingual issues.

The participating systems and evaluation results were presented in the 7th Ontology Matching workshop, that was held in Boston, MA, US [22], [7]. More information on OAEI can be found at http://oaei.ontologymatching.org/.

6.1.2. Semantics for weighted correspondences

Alignment correspondences are often assigned a weight or confidence factor by matchers. Nonetheless, few semantic accounts have been given so far for such weights. We have proposed a formal semantics for weighted correspondences between different ontologies. It is based on a classificational interpretation of correspondences: if o and o' are two ontologies used to classify a common set X, then alignments between o and o' are interpreted as encoding how elements of X classified in the concepts of o are re-classified in the concepts of o', and weights are interpreted as measures of how precise and complete re-classifications are. This semantics is justifiable for extensional matchers. We have proven that it is a conservative extension of a semantics of absolute correspondences, and we have provided properties that relate correspondence entailment with description logic constructors [8].

This work has been made in cooperation with Alexander Borgida (Rutgers University) and Chiara Ghidini and Luciano Serafini (Fondazione Bruno Kessler).

6.2. Data interlinking

The web of data uses semantic web technologies to publish data on the web in such a way that they can be interpreted and connected together. It is thus critical to be able to establish links between these data, both for the web of data and for the semantic web that it contributes to feed.

6.2.1. Keys and pseudo-keys detection for web datasets cleansing and interlinking

We have proposed a method for analysing web datasets based on key dependencies. For this purpose, we have adapted the classical notion of a key in relational databases to the case of RDF datasets [9], [16]. In order to better deal with web data of variable quality, we have introduced the definition of a pseudo-key. We have also provided an RDF vocabulary for representing keys and pseudo-keys and designed and implemented an algorithm for discovering them. Experimental results show that, even for a large dataset such as DBpedia, the runtime of the algorithm is still reasonable. We are currently working on two applications: data cleansing, i.e., detection of errors in RDF datasets and recovery, and datasets interlinking.

The algorithm is publicly available at https://gforge.inria.fr/projects/melinda/.

6.2.2. Data interlinking from expressive alignments

In the context of the DATALIFT project (see §7.1.1), we are developing a data interlinking module. Based on our analysis of the relationships between ontology matching and data interlinking [15], our goal is to generate data interlinking scripts from ontology alignments. For that purpose, we have integrated existing technologies within the DATALIFT platform: the Alignment API, for taking advantage of the EDOAL language and SILK, developed by Frei Universtität Berlin, for processing linking scripts. So far, we have generated SILK script from ontology alignments in order to produce links.

This work is part of the PhD of Zhengjie Fan, co-supervised with François Scharffe (LIRMM).

6.3. Ontology networks

Dealing with the semantic web, we are interested in ontology networks, i.e., sets of distributed ontologies that have to work together. One way for these systems to interact consists of exchanging queries and answers. For that reason, we pay particular attention to query systems.

6.3.1. Path queries and μ-calculus

Querying the semantic web is mainly done through the SPARQL language [18]. We designed one of its extensions, PSPARQL (Path SPARQL) which provides queries with paths of arbitrary length. We continue this work by connecting it to the work of the WAM team on static analysis of XPATH expressions. More specifically, we consdider query contrainment, i.e., determining whether, for any graph, the answers to a query are contained in those of another query. This is achieved by reducing this problem to satisfiability in the μ-calculus. In this work, RDF graphs are considered as transition systems and important fragments of RDFS and SPARQL as propositional μ-calculus formulas. It is then possible to use solvers of this logic to test query containment of SPARQL queries under RDFS and OWL schema constraints [11], with paths or under particular entailment regimes [10]. We have also implemented the proposed techniques and provided a first benchmark for query containment available under http://sparql-qc-bench.inrialpes.fr.

This work is part of the PhD of Melisachew Wudage Chekol [4], co-supervised with Nabil Layaïda (WAM).

FLOWERS Project-Team

6. New Results

6.1. Autonomous and Social Skill Learning and Development

6.1.1. Active Learning and Intrinsic Motivation

6.1.1.1. Active Learning of Inverse Models with Goal Babbling

Participants: Adrien Baranes, Pierre-Yves Oudeyer.

We have continued to elaborate and study our Self-Adaptive Goal Generation - Robust Intelligent Adaptive Curiosity (SAGG-RIAC) architecture as an intrinsically motivated goal exploration mechanism which allows active learning of inverse models in high-dimensional redundant robots. Based on active goal babbling, this allows a robot to efficiently and actively learn distributions of parameterized motor skills/policies that solve a corresponding distribution of parameterized tasks/goals. The architecture makes the robot sample actively novel parameterized tasks in the task space, based on a measure of competence progress, each of which triggers low-level goal-directed learning of the motor policy parameters that allow to solve it. For both learning and generalization, the system leverages regression techniques which allow to infer the motor policy parameters corresponding to a given novel parameterized task, and based on the previously learnt correspondences between policy and task parameters.

We have conducted experiments with high-dimensional continuous sensorimotor spaces in three different robotic setups: 1) learning the inverse kinematics in a highly-redundant robotic arm, 2) learning omnidirectional locomotion with motor primitives in a quadruped robot 1718 , 3) an arm learning to control a fishing rod with a flexible wire. We show that 1) exploration in the task space can be a lot faster than exploration in the actuator space for learning inverse models in redundant robots; 2) selecting goals maximizing competence progress creates developmental trajectories driving the robot to progressively focus on tasks of increasing complexity and is statistically significantly more efficient than selecting tasks randomly, as well as more efficient than different standard active motor babbling methods; 3) this architecture allows the robot to actively discover which parts of its task space it can learn to reach and which part it cannot.

This work was published in the journal Robotics and Autonomous Systems [22].

6.1.1.2. Exploration in Model-based Reinforcement Learning

Participants: Manuel Lopes, Tobias Lang, Marc Toussaint, Todd Hester, Peter Stone, Pierre-Yves Oudeyer.

Formal exploration approaches in model-based reinforcement learning estimate the accuracy of the currently learned model without consideration of the empirical prediction error. For example, PAC-MDP approaches such as R-MAX base their model certainty on the amount of collected data, while Bayesian approaches assume a prior over the transition dynamics. We propose extensions to such approaches which drive exploration solely based on empirical estimates of the learner's accuracy and learning progress. We provide a "sanity check" theoretical analysis, discussing the behavior of our extensions in the standard stationary finite state-action case. We then provide experimental studies demonstrating the robustness of these exploration measures in cases of non-stationary environments or where original approaches are misled by wrong domain assumptions. [46]. Furthermore, we studied how different exploration algorithms can be combine and selected at runtime. Typically the user must hand-tune exploration parameters for each different domain and/or algorithm that they are using. We introduced an algorithm called leo for learning to select among different exploration strategies on-line. This algorithm makes use of bandit-type algorithms to adaptively select exploration strategies based on the rewards received when following them. We show empirically that this method performs well across a set of five domains In contrast, for a given algorithm, no set of parameters is best across all domains. Our results demonstrate that the leo algorithm successfully learns the best exploration strategies on-line, increasing the received reward over static parameterizations of exploration and reducing the need for hand-tuning exploration parameters [42].

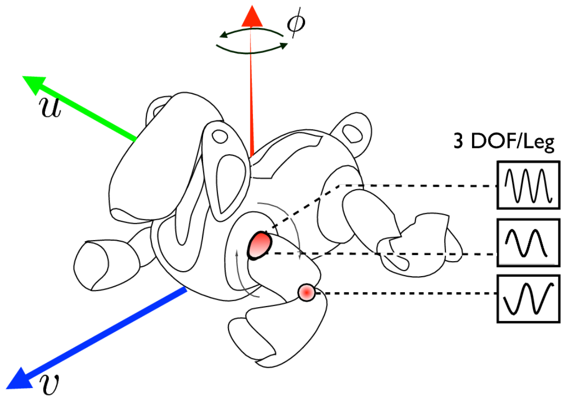


Figure 17. Experimenting SAGG-RIAC for learning an inverse model for omnidirectional locomotion of a quadruped robot. The quadruped robot is controlled using 24 dimensional motor synergies parameterized with 24 continuous values : 12 for the amplitudes and 12 others for the phases of a sinusoid tracked by each motor. Experiments consider a task space u, v, α which corresponds to the 2D position and orientation of the quadruped.

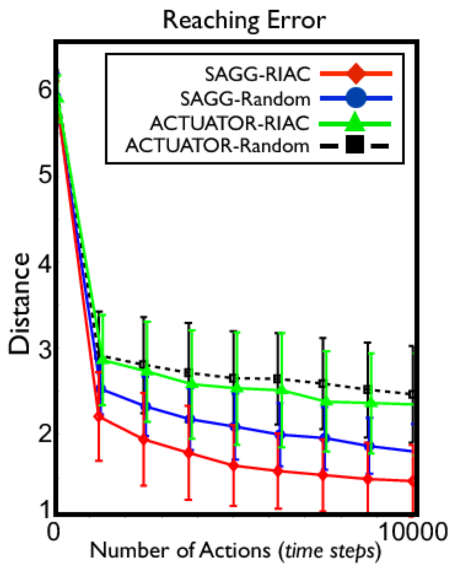
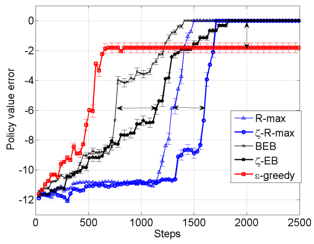
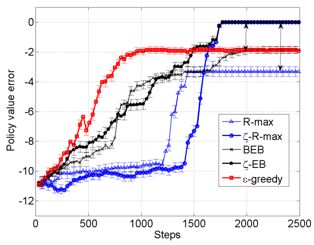


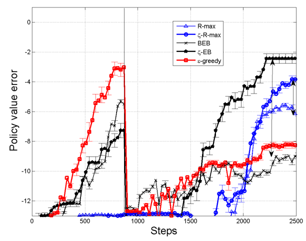
Figure 18. Evolution of the quality of the learnt inverse model for the quadruped robot experiment, depending on various exploration strategies (measured as mean error over a set of uniformly distributed goals generated independantly from learning trials).



(a) *Experiment 1—Correct Assumptions*



(b) *Experiment 2—Violated Assumptions*



(c) *Experiment 3—Change in Dynamics*

Figure 19. Experiments: (a) Like Rmax and BEB with correct assumptions, our algorithms ζ-Rmax and ζ-EB based on an empirical estimation of the learning progress converge to the optimal policy without relying on these assumptions, but take a small extra amount of time. (b) When their assumptions are violated, Rmax and BEB fail to converge, while ζ-Rmax and ζ-EB don't rely on these assumptions and again find the optimal policy. (c) In contrast to existing methods, ζ-Rmax and ζ-EB can cope with the change in transition dynamics after 900 steps and refocus their exploration.

6.1.1.3. The Strategic Student Approach for Life-Long Exploration and Learning

Participants: Manuel LOPES, Pierre-Yves OUDEYER.

We introduced and formalized a general class of learning problems for which a developmental learning strategy is shown to be optimal. This class of problems can be explained using the strategic student metaphor: a student has to learn a number of topics (or tasks) to maximize its mean score, and has to choose strategically how to allocate its time among the topics and/or which learning method to use for a given topic. We show that if the performance curves are sub-modular, then a strategy where time allocation or learning method are chosen in a developmental manner is optimal. We argue that this optimal developmental trajectory can be automatically generated by greedy maximization of learning progress. This optimal strategy amounts to creating a structured developmental exploration where typically easy tasks are first explored, and then progressively more complicated ones are explored. Furthermore, this result holds independently of the nature of the topics and the learning methods used. Then, we show an algorithm, based on multi-armed bandit techniques, that allows empirical online evaluation of learning progress and approximates the optimal solution. Finally, we show that the strategic student problem formulation allows to view in a common framework many previous approaches to active and developmental learning [47].

6.1.1.4. Active Inverse Reinforcement Learning through Generalized Binary Search

Participants: Manuel Lopes, Francisco Melo.

We contributed the first aggressive active learning algorithm for nonseparable multi-class classification. We generalize an existing active learning algorithm for binary classification [107] to the multi-class setting, and identify mild conditions under which the proposed method provably retains the main properties of the original algorithm, namely consistency and sample complexity. In particular, we show that, in the binary case, our method reduces to the original algorithm of [107]. We then contribute an extension of our method to multi-label settings, identify its main properties and discuss richer querying strategies. We conclude the paper with two illustrative application examples. The first application features a standard text-classification problem. The second application scenario features a learning from demonstration setting. In both cases we demonstrate the advantage of our active sampling approach against random sampling. We also discuss the performance of the proposed approach in terms of the derived theoretical bounds.

6.1.1.5. Towards high-dimensional and cumulative task space active exploration

Participant: Benureau Fabien.

One direction of research of the team has been on intrinsic motivation in the context of autonomous learning. Building on the PhD work of Adrien Baranes, the efforts have concentrated on creating algorithms capable to handle high-dimensional spaces and manage context with multiple tasks. The goal is for the learner to be able to autonomously create collection of reusable skills. In this context, two main research efforts have been led this year.

A typical robot is made of chains of joints. We can take advantage of the fact that joints earlier in the chain have more impact that joints further down. Given sensory feedback on the middle of the chain, an algorithm can use this information to boost learning speed and divide the learning space in subsets of smaller dimensions. We wanted to adapt this idea to high dimensional space, and specifically to the interaction with objects; a robotic arm that has already learned an inverse model of its kinematic could reuse this knowledge learn about the mapping between the position of the end-effector and the displacement of an object it is manipulating. Experiments were conducted, but they lead to the conclusion that such an approach, while effective in some specific setting, relies too heavily on a good representation of the end effector position and motion, which, in some cases, requires sensory space of higher dimension that the motor space, thus defeating the purpose. This approach was not found to be robust enough for the type of robotic context our lab is pursuing.

The SAGG-RIAC architecture is an efficient but complex architecture which implementation cannot be easily summarized in a few lines of pseudo-code. This is problematic because it reduces the ability of other research groups to implement and reuse our algorithms for their own work. An effort was started this year to create an implementation of SAGG-RIAC that would be more robust and simpler. The main idea was to use kernels rather than bins to estimate in interest in SAGG-RIAC. This approach led to very promising results, notably in its ability to handle unbounded sensory spaces. We aim at publishing the result of this work in 2013, together with a publicly available implementation of our algorithms with easy to run examples for dissemination of active learning architectures elaborated in the team. This work will also be reused in the participation of the lab into the MaCSi project.

6.1.2. Learning and optimization of motor policies

6.1.2.1. Off-Policy Actor-Critic

Participants: Thomas Degris, Martha White, Richard Sutton.

Actor–critic architectures are an interesting candidate for learning with robots: they can represent complex stochastic policies suitable for robots, they can learn online and incrementally and their per-time-step complexity scales linearly with the number of learned weights. Moreover, interesting connections have been identified in the existing literature with neuroscience. Until recently, however, practical actor–critic methods have been restricted to the on-policy setting, in which the agent learns only about the policy it is executing.

In an off-policy setting, on the other hand, an agent learns about a policy or policies different from the one it is executing. Off-policy methods have a wider range of applications and learning possibilities. Unlike on-policy methods, off-policy methods are able to, for example, learn about an optimal policy while executing an exploratory policy, learn from demonstration, and learn multiple tasks in parallel from a single sensory-motor interaction with an environment. Because of this generality, off-policy methods are of great interest in many application domains.

We have presented the first actor-critic algorithm for off-policy reinforcement learning. Our algorithm is online and incremental, and its per-time-step complexity scales linearly with the number of learned weights. We have derived an incremental, linear time and space complexity algorithm that includes eligibility traces and empirically show better or comparable performance to existing algorithms on standard reinforcement-learning benchmark problems. This work was presented by Degris et al. [38] and was reproduced independently by Saminda Abeyruwan from the University of Miami.

6.1.2.2. Auto-Actor Critic

Participant: Thomas Degris.

As mentioned above, actor–critic architectures are an interesting candidate for robots to learn new skills in unknown and changing environments. However, existing actor–critic architectures, as many machine learning algorithms, require manual tuning of different parameters to work in the real world. To be able to systematize and scale-up skill learning on a robot, learning algorithms need to be robust to their parameters. The Flowers team has been working on making existing actor–critic algorithms more robust to make them suitable to a robotic setting. Results on standard reinforcement learning benchmarks are encouraging. This work will be submitted to international conference related with reinforcement learning. Interestingly, the methods developed in this work also offer a new formalism to think about different existing themes of Flowers research such as curiosity and maturational constraints.

6.1.2.3. Relationship between Black-Box Optimization and Reinforcement Learning

Participant: Freek Stulp.

Policy improvement methods seek to optimize the parameters of a policy with respect to a utility function. There are two main approaches to performing this optimization: reinforcement learning (RL) and black-box optimization (BBO). In recent years, benchmark comparisons between RL and BBO have been made, and there has been several attempts to specify which approach works best for which types of problem classes.

We have made several contributions to this line of research by: 1) Defining four algorithmic properties that further clarify the relationship between RL and BBO. 2) Showing how the derivation of ever more powerful RL algorithms displays a trend towards BBO. 3) Continuing this trend by applying two modifications to the state-of-the-art PI^2 algorithm, which yields an algorithm we denote PI^{BB}. We show that PI^{BB} is a BBO algorithm, and, more specifically, that it is a special case of the state-of-the-art CMAES algorithm. 4) Demonstrating that the simpler PI^{BB} achieves similar or better performance than PI^2 on several evaluation tasks. 5) Analyzing why BBO outperforms RL on these tasks. These contributions have been published on HAL [69], and have been submitted to JMLR.

This work has also resulted in the novel PI^2-CMA, PI^2-CMAES algorithms, which are presented in [63], [60], [62]

6.1.2.4. Reinforcement Learning with Sequences of Motion Primitives for Robust Manipulation

Participant: Freek Stulp.

Physical contact events often allow a natural decomposition of manipulation tasks into action phases and subgoals. Within the motion primitive paradigm, each action phase corresponds to a motion primitive, and the subgoals correspond to the goal parameters of these primitives. Current state-of-the-art reinforcement learning algorithms are able to efficiently and robustly optimize the parameters of motion primitives in very high-dimensional problems. These algorithms often consider only shape parameters, which determine the trajectory between the start- and end-point of the movement. In manipulation, however, it is also crucial to optimize the goal parameters, which represent the subgoals between the motion primitives. We therefore extend the policy improvement with path integrals (PI^2) algorithm to simultaneously optimize shape and goal parameters. Applying simultaneous shape and goal learning to sequences of motion primitives leads to the novel algorithm PI^2-Seq. We use our methods to address a fundamental challenge in manipulation: improving the robustness of everyday pick-and-place tasks. This work was published in IEEE Transactions on Robotics [31] and Robotics and Autonomous Systems [26].

6.1.2.5. Model-free Reinforcement Learning of Impedance Control in Stochastic Environments

Participant: Freek Stulp.

For humans and robots, variable impedance control is an essential component for ensuring robust and safe physical interaction with the environment. Humans learn to adapt their impedance to specific tasks and environments; a capability which we continually develop and improve until we are well into our twenties. We have reproduced functionally interesting aspects of learning impedance control in humans on a simulated robot platform.

As demonstrated in numerous force field tasks, humans combine two strategies to adapt their impedance to perturbations, thereby minimizing position error and energy consumption: 1) if perturbations are unpredictable, subjects increase their impedance through co-contraction; 2) if perturbations are predictable, subjects learn a feed-forward command to offset the perturbation. We show how a 7-DOF simulated robot demonstrates similar behavior with our model-free reinforcement learning algorithm , by applying deterministic and stochastic force fields to the robot's end-effector. We show the qualitative similarity between the robot and human movements.

Our results provide a biologically plausible approach to learning appropriate impedances purely from experience, without requiring a model of either body or environment dynamics. Not requiring models also facilitates autonomous development for robots, as pre-specified models cannot be provided for each environment a robot might encounter. This work was published in IEEE Transactions on Autonomous Mental Development [29].

6.1.2.6. Probabilistic optimal control: a quasimetric approach

Participants: Clément Moulin-Frier, Jacques Droulez, Steve Nguyen.

During his previous post-doc at the Laboratoire de Physiologie de la Perception et de l'Action (Collège de France, Paris), Clément Moulin-Frier joined Jacques Droulez and Steve N'Guyen to work on an alternative and original approach of probabilistic optimal control called the quasimetric. A journal paper (soon to be submitted) was written in 2012, where the authors propose a new approach for dealing with control under uncertainty.

6.1.3. Social learning and intrinsic motivation

6.1.3.1. Optimal Teaching on Sequential Decision Tasks

Participants: Manuel Lopes, Maya Cakmak.

A helpful teacher can significantly improve the learning rate of an autonomous learning agent. Teaching algorithms have been formally studied within the field of Algorithmic Teaching. These give important insights into how a teacher can select the most informative examples while teaching a new concept. However the field has so far focused purely on classification tasks. We introduced a novel method for optimally teaching sequential decision tasks. We present an algorithm that automatically selects the set of most informative demonstrations and evaluate it on several navigation tasks. Next, we present a set of human subject studies that investigate the optimality of human teaching in these tasks. We evaluate examples naturally chosen by human teachers and found that humans are generally sub-optimal. Then based on our proposed optimal teaching algorithm we try to elicit better teaching from humans. We do this by explaining the intuition of the teaching algorithm in an informal language prior to the teaching task. We found that this improves the examples elicited from human teachers on all considered tasks. This shows that a simple modification the instructions given to human teachers, has the potential of greatly improving the performance of the agent trained by the human [32].

6.1.3.2. Socially Guided Intrinsic Motivation for Skill Learning

Participants: Sao Mai Nguyen, Pierre-Yves Oudeyer.

We have explored how social interaction can bootstrap the learning of a robot for motor learning. We first studied how simple demonstrations by teachers could have a bootstrapping effect on autonomous exploration with intrinsic motivation by building a learner who uses both imitation learning and SAGG-RIAC algorithm [22], and thus designed the SGIM-D (Socially Guided Intrinsic Motivation by Demonstration) algorithm [105]. We then investigated on the reasons of this bootstrapping effect [55], to show that demonstrations by teachers can both enhance more tasks to be explored, as well as favor more easily generalized actions to be used. This analysis is generalizable for all algorithms using social guidance and goal-oriented exploration. We then proposed to build a strategic learner who can learn multiple tasks and with multiple strategies. An overview and theoretical study of multi-task, multi-strategy Strategic Learning is presented in [47]. We also forsook to build a learning algorithm for more natural interaction with the human users. We first designed the SGIM-IM algorithm so that it can determine itself when it should ask for help from the teacher while trying to explore autonomously as long as possible so as to use as little of the teacher's time as possible [54]. After tackling with the problem of how and when to learn, we also investigated an active learner who can determine who to ask for help: in the case of two teachers available, SGIM-IM can determine which strategy to adopt between autonomous exploration and learning by demonstration, and which teacher enhances most learning progress for the learner [56], and ask him for help.

While the above results have been shown in simulation environments: of a simple deterministic air hockey game (fig. 20), and a stochastic fishing experiment with a real-time physical simulator (fig. 21), we are now building the experimental setup of the fishing experiment in order to carry out the experiments with naive users.

6.1.3.3. Adaptive task execution for implicit human-robot coordination

Participants: Ievgen Perederieiev, Manuel Lopes, Freek Stulp.

We began a project which goal is to study how computational models of multi-agent systems can be applied in situations where one agent is a human. We aim at applications where robots collaborate with humans for achieving complex tasks..

A very important capability for efficient collaborative work is the mutual agreement of a task and the ability to predict the behavior of others. We address such aspect by studying methods that increase the predictability of the robot actions. An efficient motor execution becomes the one that not just optimize speed and minimizes energy but also the one that improves the reliability of the team behavior. We are studying policy gradient methods and working on policy improvement algorithms (PI^2, CEM and $CMAES$). A feasibility study will consider a simple task between a robot and a person where the goal is to coordinate the way a set of three colored buttons is pressed.

6.1.3.4. Formalizing Imitation Learning

Participants: Thomas Cederborg, Pierre-Yves Oudeyer.

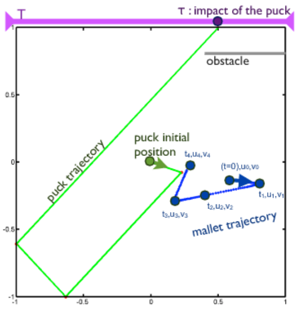


Figure 20. Illustration of SGIM-D and SGIM-IM algorithms

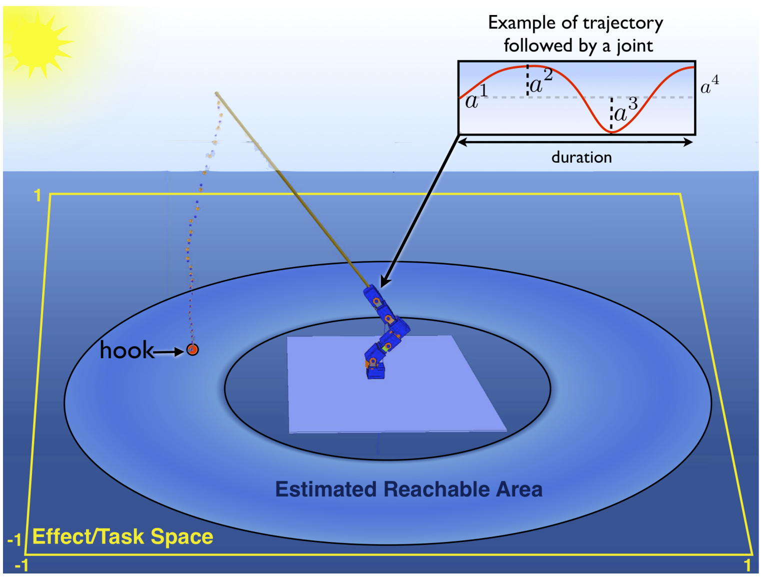


Figure 21. Illustration of SGIM-D and SGIM-IM algorithms

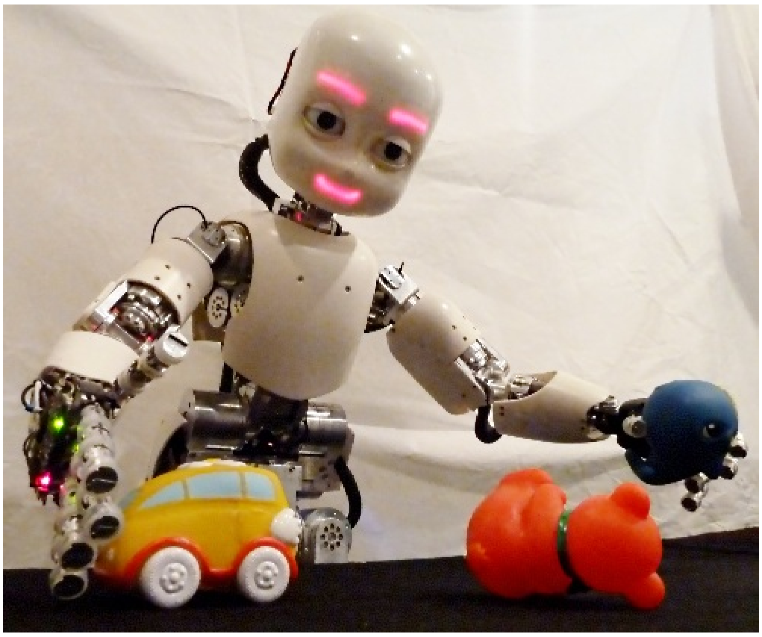


Figure 22. Illustration of SGIM-D and SGIM-IM algorithms

An original formalization of imitation learning was elaborated. Previous attempts to systematize imitation learning has been limited to categorizing different types of demonstrator goals (for example defining success in terms of the sequential joint positions of a dance, or in terms of environmental end states), and/or been limited to a smaller subset of imitation (such as learning from tele-operated demonstrations). The formalism proposed attempts to describe a large number of different types of learning algorithms using the same notation. Any type of algorithm that modifies a policy based on observations of a human, is treated as an interpretation hypothesis of this behavior. One example would be an update algorithm that updates a policy, partially based on the hypothesis that the demonstrator succeeds at demonstrations with probability 0.8, or an update algorithm that assumes that a scalar value is an accurate evaluation of an action compared to the latest seven actions. The formalism aims to give a principled way of updating these hypotheses, either rejecting some of a set of hypotheses regarding the same type of behavior, or set of parameters of an hypothesis. Any learning algorithm that modifies policy based on observations of a human that wants an agent to do something or act in some way, is describable as an interpretation hypothesis. If the learning algorithm is static, this simply corresponds to an hypothesis that is not updated based on observations. A journal article is currently being written.

6.1.4. Unsupervised learning of motor primitives

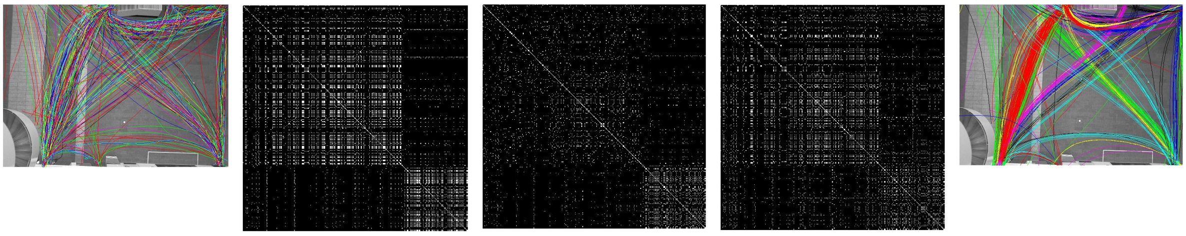
6.1.4.1. Clustering activities

Participants: Manuel Lopes, Luis Montesano.

Learning behaviors from data has applications in surveillance and monitoring systems, virtual agents and robotics among others. In our approach, ww assume that in a given unlabeled dataset of multiple behaviors, it is possible to find a latent representation in a controller space that allows to generate the different behaviors. Therefore, a natural way to group these behaviors is to search a common control system that generate them accurately.

Clustering behaviors in a latent controller space has two major challenges. First, it is necessary to select the control space that generate behaviors. This space will be parameterized by a set of features that will change for different behaviors. Usually, each controller will minimize a cost function with respect to several task features. The latent representation is in turn defined by the selected features and their corresponding weight. Second, an unknown number of such controllers is required to generate different behaviors and the grouping must be based on the ability of the controller to generate the demonstrations using a compact set of controllers.

We propose a Dirichlet Process based algorithm to cluster behaviors in a latent controller space which encodes the dynamical system generating the observed trajectories. The controller uses a potential function generated as a linear combination of features. To enforce sparsity and automatically select features for each cluster independently, we impose a conditional Laplace prior over the controller parameters. Based on this models, we derive a sparse Dirichlet Process Mixture Model (DPMM) algorithm that estimates the number of behaviors and a sparse latent controller for each of them based on a large set of features.



(a) (b) (c) (d) (e)

Figure 23. EIFPD dataset. (a) Trajectories of the EIFPD to be clustered (color is non-informative). (b-d) correspondence matrix for the 474 trajectories for the labeled ground truth, the KMeans in measurement space and the DPMM, respectively. (e) Reconstructed trajectories from the initial point using the estimated parameters of the DPMM algorithm. Due to the large number of clusters (37), colors are repeated for different clusters.

6.1.4.2. Learning the Combinatorial Structure of Demonstrated Behaviors with Inverse Feedback Control

Participants: Olivier Mangin, Pierre-Yves Oudeyer.

We have elaborated and illustrated a novel approach to learning motor skills from demonstration. This approach combines ideas from inverse feedback learning, in which actions are assumed to solve a task, and dictionary learning. In this work we introduced a new algorithm that is able to learn behaviors by assuming that the observed complex motions can be represented in a smaller dictionary of concurrent tasks. We developed an optimization formalism and show how we can learn simultaneously the dictionary and the mixture coefficients that represent each demonstration. We presented results on a idealized model where a set of potential functions represents human objectives or preferences for achieving a task in [51].

6.1.5. Maturational learning

6.1.5.1. Emergent Proximo-Distal Maturation through Adaptive Exploration

Participants: Freek Stulp, Pierre-Yves Oudeyer.

Life-long robot learning in the high-dimensional real world requires guided and structured exploration mechanisms. In this developmental context, we have investigated the use of the PI^2-CMAES episodic reinforcement learning algorithm, which is able to learn high-dimensional motor tasks through adaptive control of exploration. By studying PI^2-CMAES in a reaching task on a simulated arm, we observe two developmental properties. First, we show how PI^2-CMAES autonomously and continuously tunes the global exploration/exploitation trade-off, allowing it to re-adapt to changing tasks. Second, we show how PI^2-CMAES spontaneously self-organizes a maturational structure whilst exploring the degrees-of-freedom

(DOFs) of the motor space. In particular, it automatically demonstrates the so-called *proximo-distal maturation* observed in humans: after first freezing distal DOFs while exploring predominantly the most proximal DOF, it progressively frees exploration in DOFs along the proximo-distal body axis. These emergent properties suggest the use of PI^2-CMAES as a general tool for studying reinforcement learning of skills in life-long developmental learning contexts. This work was published in the IEEE International Conference on Developement and Learning [60].

6.1.5.2. Interaction of Maturation and Intrinsic Motivation for Developmental Learning of Motor Skills in Robots

Participants: Adrien Baranes, Pierre-Yves Oudeyer.

We have introduced an algorithmic architecture that couples adaptively models of intrinsic motivation and physiological maturation for autonomous robot learning of new motor skills. Intrinsic motivation, also called curiosity-driven learning, is a mechanism for driving exploration in active learning. Maturation denotes here mechanisms that control the evolution of certain properties of the body during development, such as the number and the spatio-temporal resolution of available sensorimotor channels. We argue that it is useful to introduce and conceptualize complex bidirectional interactions among these two mechanisms, allowing to actively control the growth of complexity in motor development in order to guide efficiently exploration and learning. We introduced a model of maturational processes, taking some functional inspiration from the myelination process in humans, and show how it can be coupled in an original and adaptive manner with the intrinsic motivation architecture SAGG-RIAC (Self-Adaptive Goal Generation - Robust Intelligent Adaptive Curiosity algorithm), creating a new system, called McSAGG-RIAC. We then conducted experiments to evaluate both qualitative and quantitative properties of these systems when applied to learning to control a high-dimensional robotic arm, as well as to learning omnidirectional locomotion in a quadruped robot equipped with motor synergies. We showed that the combination of active and maturational learning can allow to gain orders of magnitude in learning speed as well as reach better generalization performances. A journal article is currently being written.

6.1.6. Morphological computation and body intelligence

6.1.6.1. Comparative Study of the Role of Trunk in Human and Robot Balance Control

Participants: Matthieu Lapeyre [correspondant], Christophe Halgand, Jean-René Cazalet, Etienne Guillaud, Pierre-Yves Oudeyer.

Numerous studies in the field of functional motor rehabilitation were devoted to understanding the functioning of members but few are interested in the coordination of the trunk muscles and the relationship between axial and appendicular motricity which is essential in maintaining balance during travel. Acquiring new knowledge on this subject is a prerequisite in the development of new therapeutic strategies to restore motor function to the overall development of robotic orthosis that would assist the movement. Many robotic orthosis using EMG signals were unfortunately using few joints [85] and a system for controlling a multi articulated spine has not yet been developed. We propose here to use a multidisciplinary approach to define the neuro-mechanical principles where an axial system is operating in synergy with human and robot limbs.

To bring us a theoretical framework, we chose to study the reactions of the Acroban humanoid robot. Including 5 joints in the trunk, Acroban can reproduce in part the fluid movements of the human body [98] and especially to test its behavior when its trunk is held fixed or his arms are no longer used for rebalance. To disrupt postural balance in humans and robots, we have developed a low cost mobile platform (see Figure 24). This platform is made up of a broad stable support $(0.8x5m)$ mounted on a skateboard having a power of 800W. The substitution of the initial order of skate by an embedded microcontroller allows us to generate mono-axial perturbations precise intensity and duration to ensure repeatability of the disturbance. We capture movements (Optitrack 250Hz) and record the acceleration of the platform (accelerometer embedded 2kHz), the center of pressure (WiiBalanceBoard 60Hz), and electromyography (EMG).

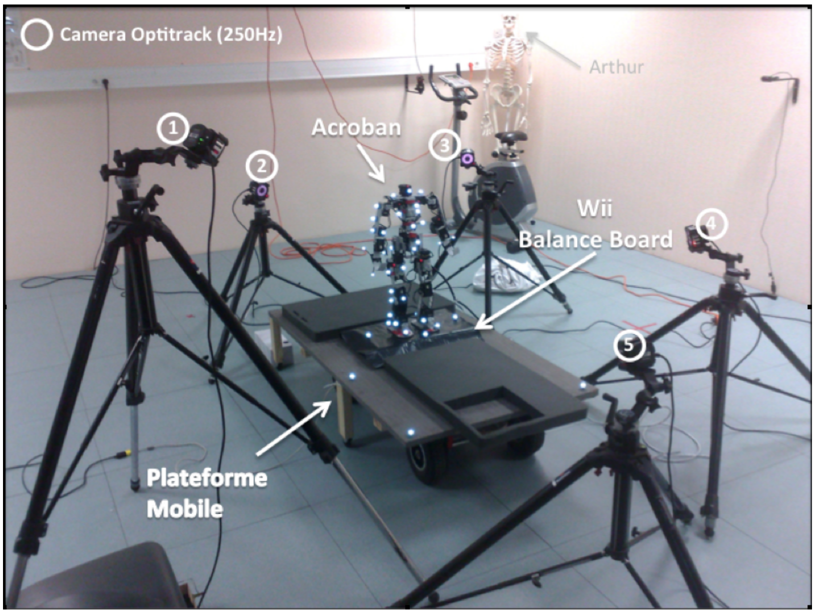


Figure 24. Experimental setup for comparative study of the role of the trunk in human and robot balance control

The experimental device (mobile platform and synchronized recordings) is operational. Preliminary experiments have allowed us to refine the profiles of disturbance on the robot Acroban. The analysis of preliminary results is in progress. Following this study, we hope to improve the modeling of the motor system in humans and robotic simulation as a basis for the development of robotic orthosis axial system. Second, the results provide a basis for improved balancing of Acroban primitives but also the development of future humanoid robots.

6.2. Autonomous and Social Perceptual Learning

6.2.1. The Impact of Human-Robot Interfaces on the Learning of Visual Objects

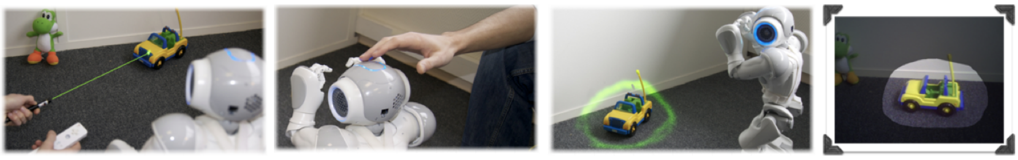
Participants: Pierre Rouanet, Pierre-Yves Oudeyer, Fabien Danieau, David Filliat.

We have continued and finalized a large-scale study of the impact of interfaces allowing non-expert users to efficiently and intuitively teach a robot to recognize new visual objects. We identified challenges that need to be addressed for real-world deployment of robots capable of learning new visual objects in interaction with everyday users. We argue that in addition to robust machine learning and computer vision methods, well-designed interfaces are crucial for learning efficiency. In particular, we argue that interfaces can be key in helping non-expert users to collect good learning examples and thus improve the performance of the overall learning system. Then, we have designed four alternative human-robot interfaces: three are based on the use of a mediating artifact (smartphone, wiimote, wiimote and laser), and one is based on natural human gestures (with a Wizard-of-Oz recognition system). These interfaces mainly vary in the kind of feedback provided to the user, allowing him to understand more or less easily what the robot is perceiving, and thus guide his way of providing training examples differently. We then evaluated the impact of these interfaces, in terms of learning efficiency, usability and user's experience, through a real world and large scale user study. In this experiment, we asked participants to teach a robot twelve different new visual objects in the context of a robotic game. This game happens in a home-like environment and was designed to motivate and engage users in an interaction where using the system was meaningful. We then analyzed results that show significant differences among

interfaces. In particular, we showed that interfaces such as the smartphone interface allows non-expert users to intuitively provide much better training examples to the robot, almost as good as expert users who are trained for this task and aware of the different visual perception and machine learning issues. We also showed that artifact-mediated teaching is significantly more efficient for robot learning, and equally good in terms of usability and user's experience, than teaching thanks to a gesture-based human-like interaction.
This work was accepted for publication in the IEEE Transactions on Robotics [28].

(a) draw the attention toward an object (b) trigger the capture (c) encircle the area of the object (d) the new learning example

Figure 25. Smartphone Interface. To make the robot collect a new learning example, users have to first draw the robot's attention toward the object they want to teach through simple gestures. Once the robot sees the object, they touch the head of the robot to trigger the capture. Then, they directly encircle the area of the image that represents the object on the screen. The selected area is then used as the new learning example. The combination of the video stream and the gestures facilitate the achievement of joint attention.



(a) draw the attention toward an object (b) trigger the capture (c) encircle the area of the object (d) the new learning example

Figure 26. Wiimote + laser pointer interface. With this interface users can draw the robot's attention with a laser pointer toward an object. The laser spot is automatically tracked by the robot. They can ensure that the robot detects the spot thanks to haptic feedback on the Wiimote. Then, they can touch the head of the robot to trigger the capture of a new learning example. Finally, they encircle the object with the laser pointer to delimit its area which will be defined as the new learning example.

6.2.2. *Curiosity-driven exploration and interactive learning of visual objects with the ICub robot*

Participants: Mai Nguyen, Serena Ivaldi, Natalia Lyubova, Alain Droniou, Damien Gerardeaux-Viret, David Filliat, Vincent Padois, Olivier Sigaud, Pierre-Yves Oudeyer.

We studied how various mechanisms for cognition and learning, such as curiosity, action selection, imitation, visual learning and interaction monitoring, can be integrated in a single embodied cognitive architecture. We have conducted an experiment with the iCub robot for active recognition of objects in 3D through curiosity-driven exploration, in which the robot can manipulate the robot or ask a human user to manipulate objects to gain information and recognise better objects (fig. 22). For this experiment carried out within the MACSi project, we address the problem of learning to recognise objects in a developmental robotics scenario. In a life-long learning perspective, a humanoid robot should be capable of improving its knowledge of objects

Figure 27. The real world environment designed to reproduce a typical living room. Many objects were added in the scene in order to make the environment cluttered.

with active perception. Our approach stems from the cognitive development of infants, exploiting active curiosity-driven manipulation to improve perceptual learning of objects. These functionalities are implemented as perception, control and active exploration modules as part of the Cognitive Architecture of the MACSi project. We integrated a bottom-up vision system based on swift feature points and motor-primitive based robot control with the SGIM-ACTS algorithm (Socially Guided Intrinsic Motivation with Active Choice of Task and Strategy as the active exploration module. SGIM-ACTS is a strategic learner who actively chooses which task to concentrate on, and which strategy is better according to this task. It thus monitors the learning progress for each strategy on all kinds of tasks, and actively interacts with the human teacher. We obtained an active object recognition approach, which exploits curiosity to guide exploration and manipulation, such that the robot can improve its knowledge of objects in an autonomous and efficient way. Experimental results show the effectiveness of our approach: the humanoid iCub is now capable of deciding autonomously which actions must be performed on objects in order to improve its knowledge, requiring a minimal assistance from its caregiver. This work constitutes the base for forthcoming research in autonomous learning of affordances.

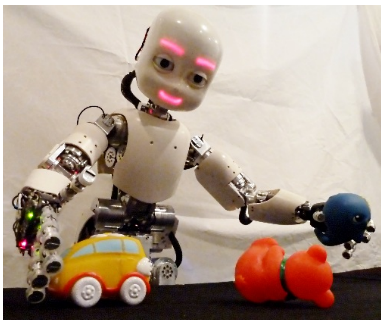


Figure 28. iCub performing curiosity-driven exploration and active recognition of visual objects in 3D

6.2.3. Discovering object concept through developmental learning

Participants: Natalia Lyubova, David Filliat.

The goal of this work is to design a visual system for a humanoid robot. Taking inspiration from child perception and following the principles of developmental robotics, the robot should detect and learn objects from interactions with people and from experiments it performs with objects, avoiding the use of image databases or of a separate training phase. In our model, all knowledge is therefore iteratively acquired from low-level features and builds up hierarchical object models, which are robust to changes in the environment, background and camera motion. In our scenario, people in front of the robot are supposed to interact with objects to encourage the robot to focus on them. We therefore assume that the robot is attracted by motion and we segment possible objects based on clustering of the optical flow. Additionally, the depth information from a Kinect is used to filter visual input, considering the constraints of the robot's working area and to refine the object contours obtained from motion segmentation.

The appearance of objects is encoded following the Bag of Visual Words approach with incremental dictionaries. We combine several complementary features to maximize the completeness of the encoded information (SURF descriptor and superpixels with associated colors) and construct pairs and triples of these features to integrate local geometry information. These features make it possible to decide if the current view has been already seen or not. A multi-view object model is then constructed by associating recognized views and views tracked during manipulations with an object.

This system is implemented on the iCub humanoid robot, which detects objects in the visual space and characterizes their appearance, their relative position and their occurrence statistics. The experiments were performed with up to ten objects; each of them was manipulated by a person during 1-2 minutes. Once the vocabulary reached a sufficient amount of knowledge, the robot was able to reliably recognize most of objects [48], [49], [43].

6.2.4. Unsupervised object categorization

Participants: Natalia Lyubova, David Filliat.

The developed unsupervised algorithm allows to identify segmented units of attention based on motion and depth information (proto-objects) into different categories such as robot hands, objects and humans.

The robot self-body category is discovered from the correlation between the proto-object positions and proprioception on the robot arms. This correlation it estimated by computing the mutual information between the changes in robot motor joints and the motion behavior of proto-objets in the visual field. The arm joints states are recorded from the robot and quantized to a vocabulary of possible arm configurations. The visual space is analyzed at the level of visual clusters that divide the perception field into regular regions. The mutual information is computed from the occurrence probabilities of the arm configurations and visual clusters.

In case of high correlation, the visual cluster is identified as a robot hand. Among the remaining proto-objects, objects are distinguished from human hands based on their quasi-static nature. Since most of objects don't move by themselves but rather are displaced by external forces, the object category is associated with regions of the visual space moving together mostly with recognized robot hands or human parts. This process make it possible to recognize the robot hands, even in case of changing appearance, and to learn to separate objects from parts of the caregivers bodies.

6.2.5. Efficient online bootstrapping of sensory representations

Participant: Alexander Gepperth.

This work [24] is a simulation-based investigation exploring a novel approach to the open-ended formation of multimodal representations in autonomous agents. In particular, we addressed here the issue of transferring (bootstrapping) features selectivities between two modalities, from a previously learned or innate reference representation to a new induced representation. We demonstrated the potential of this algorithm by several experiments with synthetic inputs modeled after a robotics scenario where multimodal object representations are bootstrapped from a (reference) representation of object affordances, focusing particularly on typical challenges in autonomous agents: absence of human supervision, changing environment statistics and limited computing power. We proposed an autonomous and local neural learning algorithm termed PROPRE (projection-prediction) that updates induced representations based on predictability: competitive advantages are given to those feature-sensitive elements that are inferable from activities in the reference representation, the key ingredient being an efficient online measure of predictability controlling learning. We verified that the proposed method is computationally efficient and stable, and that the multimodal transfer of feature selectivity is successful and robust under resource constraints. Furthermore, we successfully demonstrated robustness to noisy reference representations, non-stationary input statistics and uninformative inputs.

6.2.6. Simultaneous concept formation driven by predictability

Participants: Alexander Gepperth, Louis-Charles Caron.

This work [40] was conducted in the context of developmental learning in embodied agents who have multiple data sources (sensors) at their disposal. We developed an online learning method that simultaneously discovers meaningful concepts in the associated processing streams, extending methods such as PCA, SOM or sparse coding to the multimodal case. In addition to the avoidance of redundancies in the concepts derived from single modalities, we claim that meaningful concepts are those who have statistical relations across modalities. This is a reasonable claim because measurements by different sensors often have common cause in the external world and therefore carry correlated information. To capture such cross-modal relations while avoiding redundancy of concepts, we propose a set of interacting self-organization processes which are modulated by local predictability. To validate the fundamental applicability of the method, we conducted a plausible simulation experiment with synthetic data and found that those concepts which are predictable from other modalities successively ”grow”, i.e., become overrepresented, whereas concepts that are not predictable become systematically under-represented. We additionally explored the applicability of the developed method to real-world robotics scenarios.

6.2.7. The contribution of context: a case study of object recognition in an intelligent car

Participants: Alexander Gepperth, Michael Garcia Ortiz.

In this work [23], we explored the potential contribution of multimodal context information to object detection in an ”intelligent car”. The used car platform incorporates subsystems for the detection of objects from local visual patterns, as well as for the estimation of global scene properties (sometimes denoted scene context or just context) such as the shape of the road area or the 3D position of the ground plane. Annotated data recorded on this platform is publicly available as the a ”HRI RoadTraffic” vehicle video dataset, which formed the basis for the investigation. In order to quantify the contribution of context information, we investigated whether it can be used to infer object identity with little or no reference to local patterns of visual appearance. Using a challenging vehicle detection task based on the ”HRI RoadTraffic” dataset, we trained selected algorithms (context models) to estimate object identity from context information alone. In the course of our performance evaluations, we also analyzed the effect of typical real-world conditions (noise, high input dimensionality, environmental variation) on context model performance. As a principal result, we showed that the learning of context models is feasible with all tested algorithms, and that object identity can be estimated from context information with similar accuracy as by relying on local pattern recognition methods. We also found that the use of basis function representations [1] (also known as ”population codes” allows the simplest (and therefore most efficient) learning methods to perform best in the benchmark, suggesting that the use of context is feasible even in systems operating under strong performance constraints.

6.2.8. Co-training of context models for real-time object detection

Participant: Alexander Gepperth.

In this work[41], we developed a simple way to reduce the amount of required training data in context-based models of real- time object detection and demonstrated the feasibility of our approach in a very challenging vehicle detection scenario comprising multiple weather, environment and light conditions such as rain, snow and darkness (night). The investigation is based on a real-time detection system effectively composed of two trainable components: an exhaustive multiscale object detector (signal-driven detection), as well as a module for generating object-specific visual attention (context models) controlling the signal-driven detection process. Both parts of the system require a significant amount of ground-truth data which need to be generated by human annotation in a time-consuming and costly process. Assuming sufficient training examples for signal-based detection, we showed that a co-training step can eliminate the need for separate ground-truth data to train context models. This is achieved by directly training context models with the results of signal-driven detection. We demonstrated that this process is feasible for different qualities of signal-driven detection, and maintains the performance gains from context models. As it is by now widely accepted that signal-driven object detection can be significantly improved by context models, our method allows to train strongly improved detection systems without additional labor, and above all, cost.

6.3. Joint Learning and Development of Language and Action

6.3.1. Learning to recognize parallel motion primitives with linguistic descriptions using Non-Negative Matrix Factorization

Participants: Olivier Mangin, Pierre-Yves Oudeyer.

We have elaborated and experimented a novel approach to joint language and motor learning from demonstration. It enables discovery of a dictionary of gesture and linguistic primitives, that can be combined in parallel to represent training data as well as novel activities in the form of combinations of known gestures. These methods and the results of our experiments participate in addressing two main issues of developmental robotics: 1) symbol grounding for language learning; 2) achieving compositionality in motor-learning from demonstration, which enables re-using knowledge and thus scaling to complex tasks. In particular, we are interested in learning motor primitives active in parallel, a less explored way of combining such primitives. To address these challenges we have explored and studied the use of nonnegative matrix factorization to discover motor primitives from histogram representations of data acquired from real demonstrations of dancing movements. Initial results were presented in [99] and further results are presented in [52].

6.3.2. Curiosity-driven phonetic learning

Participants: Clément Moulin-Frier, Pierre-Yves Oudeyer.

We study how developmental phonetic learning can be guided by pure curiosity-driven exploration, also called intrinsically motivated exploration. Phonetic learning refers here to learning how to control a vocal tract to reach acoustic goals. We compare three different exploration strategies for learning the auditory-motor inverse model: random motor exploration, random goal selection with reaching, and curiosity-driven active goal selection with reaching. Using a realistic vocal tract model, we show how intrinsically motivated learning driven by competence progress can generate automatically developmental structure in both articulatory and auditory modalities, displaying patterns in line with some experimental data from infants. This work has been published in [53] and received the best paper award in computational models of development at the International Conference on Development and Learning, Epirob, San Diego, 2012.

We are now working on applying this approach to the control of a more complex articulatory synthesizer. We are interested in using the free software Praat, a powerful tool allowing to synthesize a speech signal from a trajectory in a 29-dimensional space of respiratory and oro-facial muscles. Numerous acoustic features can in turn be extracted from the synthesized sound, among which the Mel-frequency cepstral coefficients. Our hope is that a developmental robotics approach applied to a realistic articulatory model can appropriately manage the learning process of this complex mapping in high-dimensional spaces , and that observed developmental sequences can lead to interesting experimental data comparisons and predictions. In particular, using such a dynamic model controlled by muscle activity could hopefully allow to relate our results to more common speech acquisition data, in particular regarding infraphonological exploration and babbling.

6.3.3. Towards robots with teleological action and language understanding

Participants: Britta Wrede, Katharina Rohlfing, Jochen Steil, Sebastian Wrede, Jun Tani, Pierre-Yves Oudeyer.

It is generally agreed upon that in order to achieve generalizable learning capabilities of robots they need to be able to acquire compositional structures - whether in language or in action. However, in human development the capability to perceive compositional structure only evolves at a later stage. Before the capability to understand action and language in a structured, compositional way arises, infants learn in a holistic way which enables them to interact in a socially adequate way with their social and physical environment even with very limited understanding of the world, e.g. trying to take part in games without knowing the exact rules. This capability endows them with an action production advantage which elicits corrective feedback from a tutor, thus reducing the search space of possible action interpretations tremendously. In accordance with findings from developmental psychology we argue that this holistic way is in fact a teleological representation encoding a goal-directed perception of actions facilitated through communicational frames. This observation leads to a range of consequences which need to be verified and analysed in further research. We have written an article [64] where we discussed two hypotheses how this can be made accessible for action learning in robots: (1) We explored the idea that the teleological approach allows some kind of highly reduced one shot learning enabling the learner to perform a meaningful, although only partially correct action which can then be further refined through compositional approaches. (2) We discussed the possibility to transfer the concept of ”conversational frames” as recurring interaction patterns to the action domain, thus facilitating to understand the meaning of a new action. We conclude that these capabilities need to be combined with more analytical compositional learning methods in order to achieve human-like learning performance.

6.3.4. Imitation Learning and Language

Participants: Thomas Cederborg, Pierre-Yves Oudeyer.

We have studied how context-dependant imitation learning of new skills and language learning could be seen as special cases of the same mechanism. We argue that imitation learning of context-dependent skills implies complex inferences to solve what we call the ”motor Gavagai problem”, which can be viewed as a generalization of the so-called ”language Gavagai problem”. In a full symbolic framework where percepts and actions are continuous, this allows us to articulate that language may be acquired out of generic sensorimotor imitation learning mechanisms primarily dedicated at solving this motor Gavagai problem. Through the use of a computational model, we illustrate how non-linguistic and linguistic skills can be learnt concurrently, seamlessly, and without the need for symbols. We also show that there is no need to actually represent the distinction between linguistic and non-linguistic tasks, which rather appears to be in the eye of the observer of the system. This computational model leverages advanced statistical methods for imitation learning, where closed-loop motor policies are learnt from human demonstrations of behaviours that are dynamical responses to a multimodal context. A novelty here is that the multimodal context, which defines what motor policy to achieve, includes, in addition to physical objects, a human interactant which can produce acoustic waves (speech) or hand gestures (sign language). A book chapter was written and published [66] and a journal article was submitted.

6.3.5. COSMO (“Communicating about Objects using Sensory-Motor Operations”): a Bayesian modeling framework for studying speech communication and the emergence of phonological systems

Participants: Clément Moulin-Frier, Jean-Luc Schwartz, Julien Diard, Pierre Bessière.

This work began with the PhD thesis of Clement Moulin-Frier at GIPSA-Lab, Grenoble, France, supervised by Jean-Luc Schwartz (GIPSA-Lab, CNRS), Julien Diard (LPNC, CNRS) and Pierre Bessière (College de France, CNRS). A few papers were finalized during his post-doc at FLOWERS in 2012. Firstly, an international journal paper based on the PhD thesis work of Raphael Laurent (GIPSA-Lab), extending Moulin-Frier’s model, was published [25], and a commentary in *Behavioral and Brain Sciences* was accepted but not yet published [68]. Both these papers provide computational arguments based on a sensory-motor cognitive model to feed the age-old debate of motor vs. auditory theories of speech perception. Secondly, in another journal paper under the submission process, we attempt to derive some properties of phonological systems (the sound systems of human languages) from the mere properties of speech communication. We introduce a model of the cognitive architecture of a communicating agent, called COSMO (for "Communicating about Objects using Sensory-Motor Operations") that allows expressing in a probabilistic way the main theoretical trends found in the speech production and perception literature. This allows a computational comparison of these theoretical trends, helping to identify the conditions that favor the emergence of linguistic codes. We present realistic simulations of phonological system emergence showing that COSMO is able to predict the main regularities in vowel, stop consonant and syllable systems in human languages.

6.3.6. Recognizing speech in a novel accent: the Motor Theory of Speech Perception reframed

Participants: Clément Moulin-Frier, Michael Arbib.

Clément Moulin-Frier engaged this work with Michael Arbib during his 6-month visit in 2009 at the USC Brain Project, University of Southern California, Los Angeles, USA. An international journal paper is still under the revision process, in which we offer a novel computational model of foreign-accented speech adaptation, together with a thorough analysis of its implications with respect to the motor theory of speech perception.

6.3.7. Learning Simultaneously New Tasks and Feedback Models in Socially Guided Robot Learning

Participants: Manuel Lopes, Jonathan Grizou, Thomas Cederborg, Pierre-Yves Oudeyer.

We have developed a system that allows a robot to learn simultaneously new tasks and feedback models from ambiguous feedback in the context of robot learning by imitation. We have considered an inverse reinforcement learner that receives feedback from a user with an unknown and noisy protocol. The system needs to estimate simultaneously what the task is, and how the user is providing the feedback. We have further explored the problem of ambiguous protocols by considering that the words used by the teacher have an unknown relation with the action and meaning expected by the robot. This allows the system to start with a set of known symbols and learn the meaning of new ones. We have conducted human-robot interaction experiments where the user teaches a robot new tasks using natural speech with words unknown to the robot. The robot needs to estimate simultaneously what the task is and the associated meaning of words pronounced by the user. We have computational results showing that: a) it is possible to learn the task under unknown and noisy feedback, b) it is possible to reuse the acquired knowledge for learning new tasks and c) even in the presence of a known feedback, the use of extra unknown feedback signals while learning improves learning efficiency and robustness to mistakes. This algorithm has been applied on discrete and continuous problems and tested in a real world experiment using spoken words as feedback signals. A article to be submitted to a journal is currently being written.

6.3.8. Active Learning for Teaching a Robot Grounded Relational Symbols

Participants: Johannes Kulick, Tobias Lang, Marc Toussaint, Manuel Lopes.

The present work investigates an interactive teaching scenario, where a human aims to teach the robot symbols that abstract geometric (relational) features of objects. There are multiple motivations for this scenario: First, state-of-the-art methods for relational Reinforcement Learning demonstrated that we can successfully learn abstracting and well-generalizing probabilistic relational models and use them for goal-directed object manipulation. However, these methods rely on given grounded action and state symbols and raise the classical question Where do the symbols come from? Second, existing research on learning from human-robot interaction has focused mostly on the motion level (e.g., imitation learning). However, if the goal of teaching is to enable the robot to autonomously solve sequential manipulation tasks in a goal-directed manner, the human should have the possibility to teach the relevant abstractions to describe the task and let the robot eventually leverage powerful relational RL methods (see Figure 29). We formalize human-robot teaching of grounded symbols as an Active Learning problem, where the robot actively generates geometric situations that maximize his information gain about the symbol to be learnt. We demonstrate that the learned symbols can be used in a relational RL framework for the robot to learn probabilistic relational rules and use them to solve object manipulation tasks in a goal-directed manner. [44].

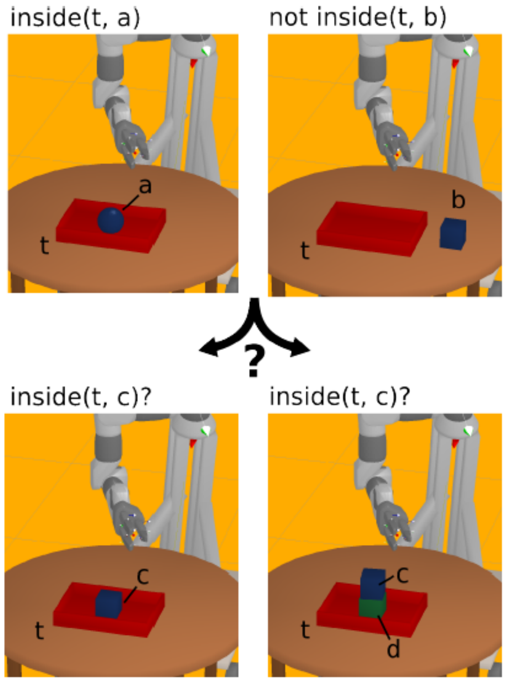


Figure 29. Active learning of symbol descriptions on a real world robot.

6.3.9. Multimodal Conversational Interaction with a Humanoid Robot

Participants: Adam Csapo, Emer Gilmartin, Jonathan Grizou, JingGuang Han, Raveesh Meena, Dimitra Anastasiou, Kristiina Jokinen, Graham Wilcock.

The paper presents a multimodal conversational interaction system for the Nao humanoid robot. The system was developed at the 8th International Summer Workshop on Multi-modal Interfaces, Metz, 2012. We implemented WikiTalk, an existing spoken dialog system for open-domain conversations, on Nao. This greatly extended the robot's interaction capabilities by enabling Nao to talk about an unlimited range of topics. In addition to speech interaction, we developed a wide range of multimodal interactive behaviours by the robot, including face- tracking, nodding, communicative gesturing, proximity detection and tactile interrupts. We made video recordings of user interactions and used questionnaires to evaluate the system. We further extended the robot's capabilities by linking Nao with Kinect. This work was presented in [34].

6.4. Other applications

6.4.1. Real-time Reaction-Diffusion Simulation: a Machine Learning Technique

Participants: Thomas Degris, Nejib Zemzemi.

Carmen is an Inria team working on modeling the electrical activity of the human heart. Their models are mainly based on reaction-diffusion equations. These methods are expansive in terms of computational costs which limits their use in practice. More specifically, some recent chirurgical intervention techniques on the heart (atrial ablation) requires to identify the source of the electrical wave. Finding such sources requires an optimization procedure. Using classical methods, this procedure is very heavy computationally.

In this project, our goal is to reduce the computational cost using supervised learning techniques. The idea is to replace the incremental resolution of partial differential equations by more suitable data structures for real-time running. Starting from data generated by simulating different excitations scenari on a human atria, this data is afterwords used as a training data set for machine learning algorithms. This approach will allow a faster optimization procedure.

This work is in collaboration with Nejib Zemzemi from the Inria Carmen team. This project is in preliminary steps.

GRAPHIK Project-Team

6. New Results

6.1. Ontology-Based Query Answering with Existential Rules

Participants: Jean-François Baget, Mélanie König, Michel Leclère, Marie-Laure Mugnier, Michaël Thomazo.

Note that for this section, as well as all sections in New Results, participants are given in alphabetical order.

In collaboration with: Sebastian Rudolph (Karlsruhe Institute of Technology)

We have pursued the work on the existential rule framework in the context of Ontology-Based Query Answering. See the 2011 activity report for details on this framework also known as Datalog+/-. The ontology-based query answering problem consists of querying data while taking into account inferences enabled by an ontology (described by existential rules in our case).

From 2009 to 2011, we mainly investigated decidability and complexity issues. In 2012, while still interested in deepening decidability and complexity results, we tackled the next step: algorithms. Our aim is to develop algorithms with good theoretical properties (at least they should run in "the good worst-case complexity class") and with good performance in practice. There are two main ways of processing rules, namely forward chaining and backward chaining. In forward chaining, rules are applied to enrich the initial facts and query answering is solved by evaluating the query against the "saturated" facts (as in a classical database system). When it is finite, the backward chaining process can be divided into two steps: first, the query is rewritten into a first-order query (typically a union of conjunctive queries) using the rules; then the rewritten query is evaluated against the initial facts (again, as in a classical database system).

6.1.1. Forward Chaining Algorithms

Considering the expressive class of greedy bounded-treewidth set of rules (in short *gbts*), which we defined in 2011, we have designed a query answering algorithm which has several advantages over 2011 algorithm, while staying optimal with respect to worst-case combined and data complexities.

1. It is much more implementable (previous algorithm was using an oracle).
2. It is generic in the sense that it works for any class of rules that fulfills the gbts property, but it can also be easily specialized for specific gbts subclasses whith lower complexities, such as frontier-guarded or guarded rules, in such a way that it runs in the good complexity class.
3. It allows for separation between offline and online processing steps: the knowledge base can be compiled independently from queries, which are evaluated against the compiled form.

One of the lightweight description logics used for ontology-based query answering is $\mathcal{EL}$. We designed a subclass of existential rules that covers $\mathcal{EL}$ with the same complexity of reasoning, while allowing for any predicate arity and some cycles on variables. We also added complex role inclusions like transitivity and right/left identity rules to enhance expressivity, while staying polynomial in data complexity and generalizing existing results.

- *Results published in [36], [37] and [32] (invited conference). See also our research report [49] for a longer version.*
- *A journal version extending the papers at IJCAI 2011 and KR 2012 is in preparation, to be submitted to a major artificial intelligence journal.*

6.1.2. Backward Chaining Algorithms

We consider query rewriting techniques that output a union of conjunctive queries, which we see as a set of conjunctive queries. More specifically, only the most general elements of this set need to be kept in the output. We first proved that all sound and complete query rewriting algorithms necessarily produce the same result (up to redundancy) when restricted to their most general elements. It follows that comparing existing algorithms with respect to the size of the produced query is pointless.

Existing query rewriting algorithms accept only specific classes of existential rules (mainly corresponding to the translation of some lightweight description logics). We designed an algorithm that accept as input any set of existential rules and stops if this set of rules fufills so-called *fus* property (meaning that the set of most general rewritings of any initial conjunctive query is finite). This algorithm has been implemented and first experimentations have been led on rule bases obtained by translating description logic bases.

- *Results published in [31] (best paper price)*

6.1.3. Querying Optimization (Work in Progress)

Our current work aims at improving previous algorithms, in particular: the online querying step in the gbts algorithm; the query rewriting algorithm, by avoiding generating several times equivalent rewritings; for specific subclasses, query rewriting into a set of so-called semi-conjunctive queries instead of conjunctive queries, which reduces the size of the output query.

6.2. Reasoning with Imperfect Information and Priorities

Participants: Madalina Croitoru, Jérôme Fortin, Souhila Kaci, Tjitze Rienstra, Rallou Thomopoulos.

In collaboration with: Joël Abecassis (IATE/INRA), Patrice Buche (IATE/INRA), Nir Oren (Univ. of Aberdeen, Scotland), Leon van der Torre (University of Luxembourg) and Nouredine Tamani (post-doc IATE).

This year, we mainly investigated decision support based on argumentations systems and preferences, either in relation with application needs in agronomy or on more fundamental aspects.

6.2.1. Argumentation for Decision Making in Agronomy

Historically, scientific investigations in this axis are guided by applications of our partners in agronomy (IATE laboratory). Part of the work has consisted of analyzing the proposed applications and the techniques they require in order to select appropriate applications with respect to our team project.

In the context of the EcoBioCap project (see Sect. 8.2), the different stakeholders have expressed conflicting preferences for the packaging quality. However, when discussing with domain experts they have raised the need for a tool which allows them to highlight a conflict and see the reasons behind it. In order to achieve this goal two steps were taken. First we have instantiated a popular logical argumentation framework (ASPIC+) with a simple preference logic. This allowed the different experts to express arguments about their preferences. We can then extract maximal consistent subsets of preferences by the means of extensions.

- *This work was performed in collaboration with the University of Aberdeen (Dr. Nir Oren) and the results were published and presented at the COMMA conference [24].*

Second, a negotiation phase was introduced to the previously described system in order for the domain experts to refine and extend their preferences. This tool was the aim of the master thesis of Patricio Mosse.

- *This work was published and presented at the Effost conference [23], based upon Patricio Mosse's Master Thesis [48]. A detailed journal article reporting on the two steps is under preparation and will be submitted beginning 2013.*

Let us mention additional results related to the applications in agronomy on semi-automatic data extraction from web data (tables) [39], [40], [41], data reliability, and the representation and flexible querying of imprecise data with fuzzy sets [42], [15]. These investigations are complementary to the above mentioned results on argumentation and generally relate to other aspects in the same applicative projects.

6.2.2. Conditional Acceptance Functions

Dung-style abstract argumentation theory centers on argumentation frameworks and acceptance functions. The latter take as input a framework and return sets of labelings. A labeling assigns “in”, “out” or “undecided” to each arguments. Arguments having “in” assignment are acceptable arguments. This methodology however assumes full awareness of the arguments relevant to the evaluation. There are two reasons why this is not satisfactory. Firstly, full awareness is, in general, not a realistic assumption. Second, frameworks have explanatory power, which allows us to reason abductively or counterfactually, but this is lost under the usual semantics. To recover this aspect, we generalized conventional acceptance, and we present the concept of a conditional acceptance function which copes with the dynamics of argumentation frameworks.

- *Results published in [28].*

6.2.3. Foundational Aspects of Preferences

Preferences are the backbone of various fields as they naturally arise and play an important role in many real-life decisions. Preferences are fundamental in scientific research frameworks as well as applications. One of the main problems an individual faces when expressing her preferences lies in the number of variables (or attributes or criteria) that she takes into account to evaluate the different outcomes. Indeed, the number of outcomes increases exponentially with the number of variables. Moreover, due to their cognitive limitation, individuals are generally not willing to compare all possible pairs of outcomes or evaluate them individually. These facts have an unfortunate consequence that any preference representation language that is based on the direct assessment of individual preferences over the complete set of outcomes is simply infeasible.
Fortunately, individuals can abstract their preferences. More specifically, instead of providing preferences over outcomes (by pairwise comparison or individual evaluation), they generally express preferences over partial descriptions of outcomes. Often such statements take the form of qualitative comparative preference statements e.g., "I like London more than Paris" and "prefer tea to coffee". Conditional logics aim at representing such partial descriptions of individual preferences which we refer to as comparative preference statements. They use different completion principles in order to compute a preference relation induced by a set of preference statements. In particular they use various more or less strong semantics to interpret comparative preference statements. So far the main objective in artificial intelligence has been to rank-order the set of outcomes given a set of comparative preference statements and one or several semantics. We addressed this problem from a different angle. We considered a set of postulates studied in preference logics and non-monotonic reasoning which formalize intuition one may have regarding the behavior of preference statements. We analyzed the behavior of the different semantics w.r.t. these postulates. Our analysis gives a complete picture of the behavior of our (five) semantics.

In the last decade, AI researchers have pointed out the existence of two types of information: positive information and negative information. This distinction has also been asserted in cognitive psychology. Distinguishing between these two types of information may be useful in both knowledge and preference representation. In the first case, one distinguishes between situations which are not impossible because they are not ruled out by the available knowledge, and what is possible for sure. In the second case, one distinguishes between what is not rejected and what is really desired. Besides it has been shown that possibility theory is a convenient tool to model and distinguish between these two types of information. Knowledge/Preference representation languages have also been extended to cope with this particular kind of information. Nevertheless despite solid theoretical advances in this topic, the crucial question of "which reading (negative or positive) one should have" remains a real bottleneck. We focused on comparative statements and presented a set of postulates describing different situations one may encounter. We provided a representation theorem describing which sets of postulates are satisfied by which kind of information (negative or positive) and conversely. One can then decide which reading to apply depending on which postulates she privileges.

- *Results published in [29] and [30].*

6.2.4. Argumentation for Inconsistency-Tolerant Query Answering (Work in Progress)

Argumentation allows to encode by the means of extensions maximal subsets of the knowledge base which are consistent (given the logic chosen). We are currently investigating the link between different argumentation extensions and the notion of a maximal repair as introduced by [51], [50] in the context of the positive existential subset of first order logic we are mainly working with. We are then interested in comparing the semantics proposed in the literature for query answering with inconsistent knowledge bases and argumentation reasoning paradigms. This study has been performed jointly with the University of Luxembourg during a research visit during end of November. We plan to submit our results at a conference beginning January.

6.3. Semantic Data Integration

Participants: Michel Chein, Madalina Croitoru, Léa Guizol, Michel Leclère, Rallou Thomopoulos.

It often happens that different references (i.e. data descriptions), possibly coming from heterogeneous data sources, concern the same real world entity. In such cases, it is necessary: (i) to detect whether different data descriptions really refer to the same real world entity and (ii) to fuse them into a unique representation. Since the seminal paper [52], this issue has been been studied under various names:“record linking”, “entity resolution”,“reference resolution”, ”de-duplication”, “object identification”, “data reconciliation”, etc., mostly in databases (cf. the bibliography by William E. Winckler [1]). It has become one of the major challenges in the Web of Data, where the objective is to link data published on the web and to process them as a single distributed database. Most entity resolution methods are based on classification techniques; Fatiha Saïs, Nathalie Pernelle and Marie-Christine Rousset proposed the first logical approach [53]. Many experiments on public data are underway, in France (cf. DataLift[2] and ISIDORE[3] projects) or internationally (e.g., VIAF project[4] led by OCLC[5], whose aim is to interconnect authority files coming from 18 national organizations).

Three years ago, we began a collaboration with ABES (National Bibliographic Agency for Universities, which takes part in the VIAF project). The aim of this collaboration is to enable the publication of ABES metadata based on the Web of Data and to provide an identification service dedicated to bibliographic notices. ABES bibliographic bases, and more generally document metadata bases, appear to be a privileged application domain for the representation and reasoning formalisms developed by the team. This work has an interdisciplinary dimension, as it also requires experts in the Library and Information Science domain. We think that a logical approach is able to provide a generic solution for entity resolution in document metedata bases, even though it is generally admitted in Library and Information Science that *“there is no single paradigmatic author name disambiguation task—each bibliographic database, each digital library, and each collection of publications, has its own unique set of problems and issues”* [54].

6.3.1. Implementation of an Entity Identification Service

Last year, we have developed a method and a prototype to perform entity resolution between on one hand the authors of a new bibliographic notice, and, on the other the domain experts of an authority catalog (and namely the Sudoc catalogue from the ABES agency). The prototype providing this service has been implemented on top of Cogui and experiments have been led in the context of the SudocAd project (jointly conducted by ABES and GraphIK). This work has been continued this year on the following issues as part of the Qualinca project:

- *generalizing the developed method with the aim to define a generic combined (numerical/logical) framework for entity resolution. This work is reported in the research report [44] that we plan to submit to a conference in January.*
- *Defining evaluation measures of the quality of an entity resolution tool. This work is still on-going.*

6.3.2. Quality of Document Catalogs

The SudocAd project showed the feasability and pertinence of a mixed approach for data interlinking problems. It also showed the immediate necessity of taking into account the existence of human errors already present in document catalogues. This led us to propose Qualinca, an ANR Contint project, accepted beginning 2012 and started in April 2012. The partners include two major actors in the document catalogues field: ABES and INA, as well as three academic research groups.

In this context we currently investigate a formal approach to the notion of a "key" in the web of data. Our immediate objective is to define the notion of a discovered key used then in order to evaluate the quality of data inter linking of a meta data catalogue.

[1] http://www.hcp.med.harvard.edu/statistics/survey-soft/docs/WinklerReclinkRef.pdf
[2] DataLift, http://datalift.org/
[3] ISIDORE, http://www.rechercheisidore.fr/
[4] The Virtual International Authority File, http://www.oclc.org/research/activities/viaf/
[5] Online Computer Library Center, http://www.oclc.org

We also study the methodology of linking error detection and fixing based on a partitioning (clustering) method on authors of bibliographic records. This study is part of the PhD thesis of LÃ©a Guizol (jointly funded by GraphIK and ABES). The above mentioned methodology is based on a set of criteria which will allow us to cluster "similar" authors together. Each criterion represents a point of view on the author: name, publication time span, publication domain etc. The first challenge consists of defining for each of such view points the respective criteria. The second challenge is to propose an aggregation semantics of such criteria which is well adapted for the problem at hand.

- *The methodology of using such clustering techniques for this problem has been published in [25]. A certain number of criteria have already been implemented and different partitioning semantics proposed. We are currently evaluating these on the ABES data.*

6.3.3. Multi Agent Knowledge Allocation

The assumption behind semantic data integration and querying is that different agents accessing the integrated data repository will have equal interest in the querying results. This is not always true in a data sensitive scenario where the knowledge provider might want to allocate the query answers to the agents based on their valuations. Furthermore, the agents might want some information exclusively (and thus offer a valuation that allows it) while others might want it shared. To this end we have proposed a new mechanism of allocation of query answers inspired from combinatorial auctions. We have defined the newly introduced scenario of Multi Agent Knowledge Allocation and proposed a graph based method, inspired on network flows, for solving it.

- *These results were published in [26] and [35]. We are currently investigating the mechanism design aspects of such valuations in collaboration with the University of Athens (Dr. Iannis Vetsikas).*

IMAGINE Team

6. New Results

6.1. Introduction

We are developing user-centred, knowledge-based models in three main domains: shape, motion and narrative design, leading us to three research axes. The fourth one is the combination of these models with intuitive interaction tools, in order to set up interactive creative environments dedicated to specific categories of content. The following sections describe our activities in 2012 for each axis.

6.2. High level model for shapes

Scientist in charge: Stefanie Hahmann

Other permanent researchers: Marie-Paule Cani, Jean-Claude Léon, Damien Rohmer.

6.2.1. Implicit surface modeling

Participants: Adrien Bernhardt, Marie-Paule Cani, Maxime Quiblier, Cédric Zanni.

Implicit surfaces are an appealing representation for free-form, volumetric shapes. In addition to being able to represent shapes of arbitrary topological genius, they have the ability to be constructed by successively blending different components, which eases interactive modeling.

In collaboration with a researcher in formal computation, Evelyne Hubert, we improved and extended the analytical methods for computing closed form solutions for convolution surfaces [6].

Within Cédric Zanni's PhD we proposed a method based on anisotropic, surface Gabor noise, for generating procedural details on skeleton-based implicit surfaces, see Figure 4 (left). The surfaces enhanced with details can still be smoothly blended, with a natural transition between the details they carry [19].

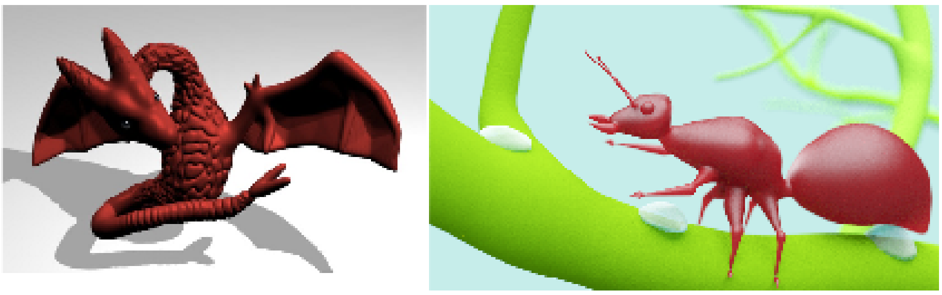


Figure 4. Left: Dragon model showing the variety of details that can be generated. Computation time was less than 2 minutes. Right: Shape obtained by the use of scale-invariant integral surfaces.

We also developed an extension to convolution surfaces, so-called scale-invariant integral surfaces, see Figure 4 (right). Thanks to blending properties that are scale invariant these surfaces have three major advantages: the radius of the surface around a skeleton can be explicitly controlled, shapes generated in blending regions are self-similar regardless of the scale of the model, and thin shape components are not smoothed-out anymore when blended into larger ones. This work has been presented at AFIG2012 [23] and submitted for international publication.

Lastly, in collaboration with Loic Barthe in Toulouse, we contributed to a new blending operator, gradient blending, which enables us to blend implicit shapes not only in function of the field values but also of their gradients. This solves a number long standing problems in implicit modeling: we can generate bulge-free blending, ensure that the topological genius of the blended shape remains the one of the union of the input one, and avoid the blur of small details. A paper has been accepted for publication in ACM ToG [4].

6.2.2. Developable surfaces

Participants: Rémi Brouet, Marie-Paule Cani, Stefanie Hahmann, Damien Rohmer.

Figure 5. Design preserving garment transfer of a multi-layer outfit from a woman to a young girl. Middle: Automatically graded patterns shown to scale. Right: The zoomed-in source and target patterns for the back panel highlight the subtle changes in shape

A developable surface is a surface, which can be unfolded (developed) into a plane without stretching or tearing. Because of this property, developable surfaces lead to a variety of applications in manufacturing with materials that are not amenable to stretching (leather for shoes or hand bags, skins of aircrafts, sails). In computer graphics developable surfaces are very popular to model, simulate or animate clothes or folded papers in virtual environments.

In collaboration with Alla Sheffer (University of British Columbia, Canada visiting Inria) we developed a fully automatic method for design-preserving transfer of garments between characters with different body shapes. The method is able to generate design-preserving versions of existing garments for target characters whose proportions and body shape significantly differ from those of the source. The work has been presented at SIGGRAPH 2012 [1].

Folded paper exhibits very characteristic shapes, due to the presence of sharp folds and to exact isometry with a given planar pattern. In the past we proposed a purely geometric solution to generate static folded paper geometry from a 2D pattern and a 3D placement of its contour curve. Current research focuss on the interactive manipulation of the folded surface without the strong requirement of starting by an initial contour curve, but using sparser positional constraints on the surface.

Damien Rohmer joined in 2012 the Hevea project: this is a project in collaboration between Vincent Borrelli (Institut Camille Jordan, Lyon), Boris Thibert (MGMI, LJK Grenoble) and Francis Lazarus (Gipsa Lab, Grenoble) focussed on the generation and visualisation of the flat torus. The flat torus is a mathematical smooth surface with the topology of a torus but having locally the metric of the plane. In other word, this

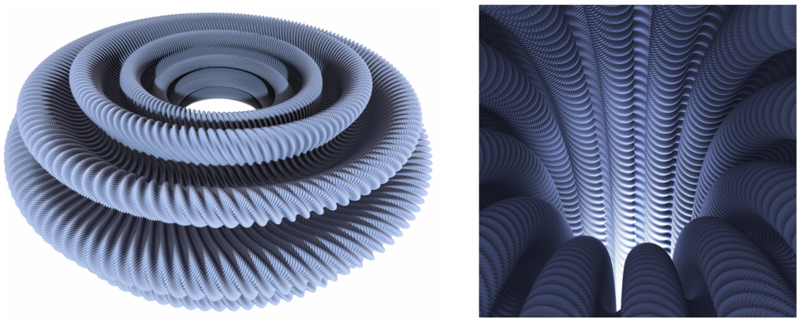


Figure 6. The first representation of a flat torus.

is a developable torus. So far, no representation of such object had ever being made. In 2012, based on a convex integration algorithm generating coherent wrinkles on the torus called *corrugations*, we generated the first representation of such object that is both C^1 while being fractal as the number of wrinkles has to tend to infinity to converge toward true developability. The rendering made by Damien Rohmer has been used for the cover image of Proceedings of the National Academy of Sciences (PNAS) (http://www.pnas.org/content/109/19.cover-expansion).

6.2.3. Parametric surfaces

Participant: Stefanie Hahmann.

We are developing new smooth parametric surface models defined on irregular quad meshes. They are in fact a powerful alternative to singularly parameterized tensor product surfaces since they combine the advantages of both, the arbitrary topology of quad meshes and the smoothness of the tensor product patches.

In collaboration with G.-P. Bonneau (Maverick team) several parametric triangular surface models for arbitrary topologies have been published in the past (CAGD, IEEE TVCG and ACM ToG). A new tensor product spline surface model has been developed this year. It solves the problem of defining a G^1-continuous surface interpolating the vertices of an irregular quad mesh with low degree polynomial tensor product patches. It further aims to produce shapes of very high visual quality while reducing the number of control points. A comparison with existing methods and a journal paper are in preparation.

6.2.4. Fibrous structures

Participant: Damien Rohmer.

Due to anisotropy, fibrous structures may exhibit complex deformation properties. These properties are of main interest to understand the behavior of some human organs such as the heart.

In collaboration with Grant Gullberg, Archontis Giannakidis from Lawrence Berkeley Laboratory, and Alexander Veress from University of Washington we developped a new visualization of heart defects based on the fibrous structure organization. Combining 3D visualization with the fiber structure analysis may help to detect heart defects such as cardiac Hypertrophy. This work as been published as a book chapter [29].

6.2.5. Virtual Prototypes

Participants: Flavien Boussuge, Francois Faure, Stefanie Hahmann, Jean-Claude Léon.

In the context of virtual prototyping (process of product development involving CAD/CAE software), a DMU (digital mock up) is the container of all the components of a 3D virtual product that be used during design and simulations.

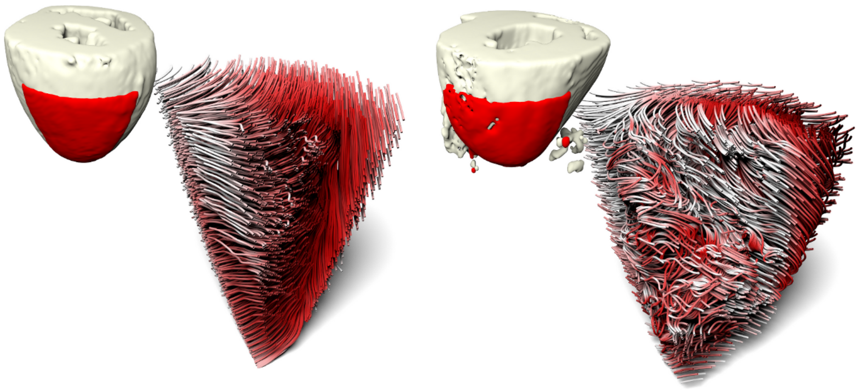


Figure 7. Visualization of fibrous structure in the heart for a normal case (left) and a defect heart (right).

Herein geometric interfaces, i.e. the imprint of a component onto each of its neighboring components, must be taken into account to generate simulation models. Indeed, a DMU does not contain these geometric interfaces. However, extensive use of CAD assemblies has led to increasingly complex DMUs with up to hundreds of thousands of components. The detection and generation of the geometric interfaces between all components with existing software is a very tedious task, which may require hours or days of user-interaction or is even not possible. As part of the ANR project ROMMA in collaboration with Georges-Pierre Bonneau and Francois Jourdes from the Maverick team, we developed a new method to rapidly detect and precisely describe the geometry of interfaces in highly complex assemblies [20].

Within the PhD of Flavien Boussuge, we take advantage of these interfaces to focus on the generation of mixed dimensional models from enriched DMUs for FE analysis of structural assemblies. The goal is to provide a methodology and operators for transforming geometries of complex assemblies so that they are directly usable for FE mesh generation. A first contribution to assembly model preparation for simulation has been presented at ECT12 [11]. Herein, a model preparation methodology has been proposed that addresses the shape transformation categories specific to assemblies. Current and future research includes the generation of construction graphs of volume models that contribute to idealization operators. These algorithms take the simulation objectives into account as part of the proposed methodology.

Another important issue connected to geometry transformation of assemblies and construction graphs of volume models relates to the global as well as partial symmetries of components and assemblies. Here, symmetry analysis is applied to B-Rep NURBS models and must be obtained within the tolerance of a geometric modeler, which differs rather significantly from approximate symmetries extracted from meshes. The symmetry analysis helps structuring the construction graphs of volume models to take into account repetitive locations of primitives. Also, symmetry properties combine with functional annotations of components to enhance their search and retrieval[16].

6.3. Models for real-time motion synthesis

Scientist in charge: François Faure

Other permanent researchers: Marie-Paule Cani, Damien Rohmer, Rémi Ronfard.

6.3.1. Interactive manipulation of folded paper surfaces

Participant: Damien Rohmer.

Although physically-based simulation has become very popular to model deformable surfaces such as cloth it is still not applicable to generate animations of creased paper. Due to the stiffness of this uncompressible material and to the complex changes of its mechanical behavior during creasing. As a result, this standard material in every-day life almost never appears in Computer Graphics applications such as movies or video games. Animating creased paper brings two main challenges: First, such surface needs to be deformed while preserving its length in every direction according to its original pattern. Secondly, sharp features, which are not commonly handled in numerical simulators, need to be generated on the surface.

With the master work of Ulysse Vimont, we developped a prototype (as seen in fig. 8) of a deformation tool enabling to interactively manipulate a virtual sheet of paper. The approach is a procedural approach based on some geometrical apriori knowledge of behavior of paper under deformation. We plan to extend this work in the next year with a new master student Camille Shreck.

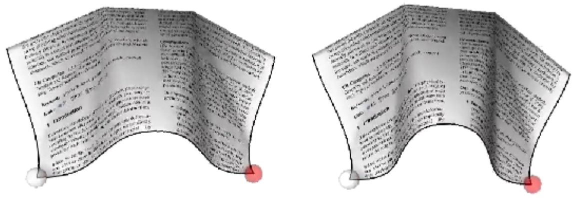


Figure 8. Example of interactive manipulation of a sheet of paper.

6.3.2. Real-time skinning deformation with contacts

Participants: Marie-Paule Cani, Damien Rohmer.

Skinning deformation based on linear blending or dual quaternion approach is a very popular technique thanks to its fast computation. However, they do not capture the complex behavior of flesh bulging and contact between body parts.

In collaboration with Loic Barthe, Rodolphe Vaillant from IRIT Toulouse, and Gael Guennebaud from LaBRI Bordeaux, we developped a skinnning deformation handling both flesh bulges and collision avoidance.

An implicit surface is first fitted onto the original mesh surface. During the animation, the mesh is deformed using a standard skinning deformation while the implicit surface follows the rigid articulation of the bone. Finaly, the mesh is projected back toward the deformed implicit surface enabling to both compensate for mesh collapse and self collision. This work has been presented in AFIG [10] conference and won the *best article* award. It has also being accepted for publication in the REFIG Journal.

6.3.3. Particle-based simulation of concrete structures

Participants: Marie Durand, François Faure.

In collaboration with the LIG and L3S-R labs, we have published results on gpu-accelerated simulation of concrete fracturation due to impacts [2], leading to a speedup factor of about 15 compared to a CPU implementation. This led us to notice that collision detection was the major bottleneck. Consequently, we investigated and published a new incremental sorting method to more efficiently cluster the particles along a Z-curve, by improving the Packed Memory Array data structure for fast updates [15], as illustrated in Figure 9 . We have proposed a new strategy to efficiently update the sorting, while maintaining a desired fill rate in each branch of the tree structure. Experiments show that our PMA can outperform a compact sorted array for up to 50% particle cell changes per time step.

6.3.4. Collision detection and response

Participant: François Faure.

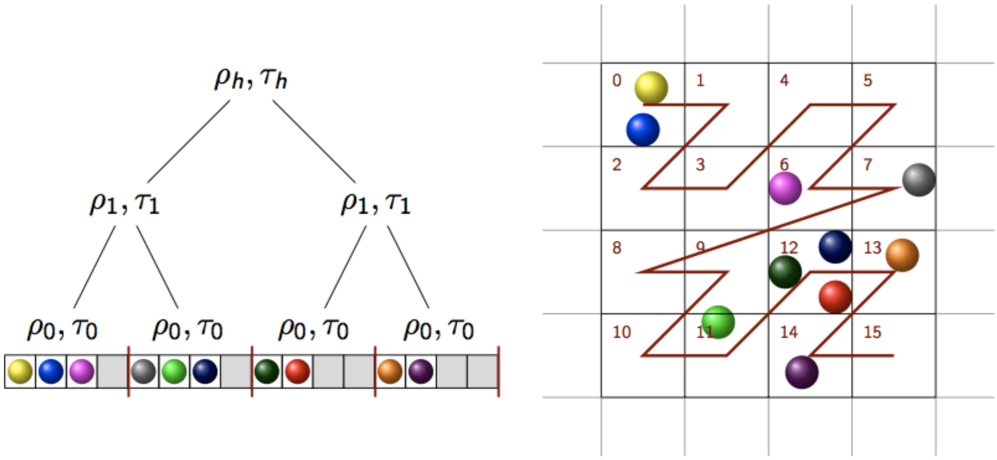


Figure 9. Z-sort using a Packed Memory Array structure. The gaps allow fast updates.

In collaboration with UBC, Vancouver, we have presented at SIGGRAPH 2012 a new method for image-based contact detection and modeling, with guaranteed precision on the intersection volume [8]. Unlike previous image-based methods, our method optimizes a nonuniform ray sampling resolution and allows precise control of the volume error. By cumulatively projecting all mesh edges into a generalized 2D texture, we construct a novel data structure, the Error Bound Polynomial Image (EBPI), which allows efficient computation of the maximum volume error as a function of ray density. Based on a precision criterion, EBPI pixels are subdivided or clustered. The rays are then cast in the projection direction according to the non-uniform resolution. The EBPI data, combined with ray-surface intersection points and normals, is also used to detect transient edges at surface intersections. This allows us to model intersection volumes at arbitrary resolution, while avoiding the geometric computation of mesh intersections. Moreover, the ray casting acceleration data structures can be reused for the generation of high quality images, as illustrated in Figure 10 .

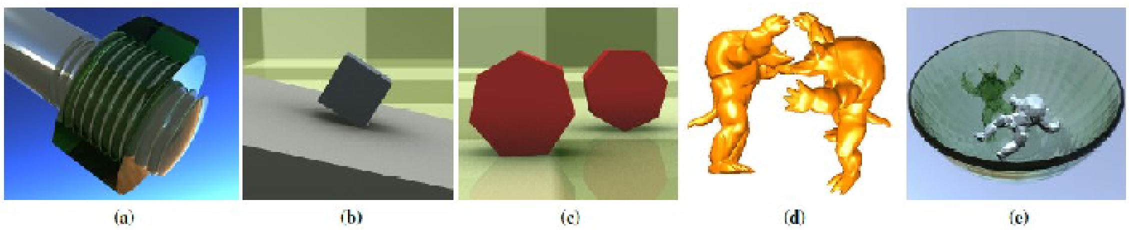


Figure 10. Examples of challenging contact scenarios handled by our method. (a) The movement of a tight fitting nut on a bolt can be simulated directly using the geometric models. (b) Very small geometric features on a flat surface can dramatically change the behavior of objects sliding on it. (c) "Ruina wheels." Two almost identical wheels have obviously different rolling behavior due to subtle features (one is slightly convex and another is slightly concave); our method can simulate this contact behavior realistically. (d) A simulation with 4:4 million triangles. (e) A snapshot of an interactive simulation with ray-traced rendering.

6.3.5. Action representation, segmentation and recognition

Participant: Remi Ronfard.

Following Daniel Weinland's PhD thesis, we published a survey of modern methods for representing, segmenting and recognizing full-body actions in video [32]. A taxonomy of methods is elaborated in that paper, where actions can be represented with local, structured or global features both in time and in space. The potential for future work in grammar-based action recognition is emphasized, with possible applications in corpus-based procedural modeling of actions.

6.3.6. Simulation software architecture

Participants: Ali-Hamadi Dicko, Guillaume Bousquet, Françcois Faure.

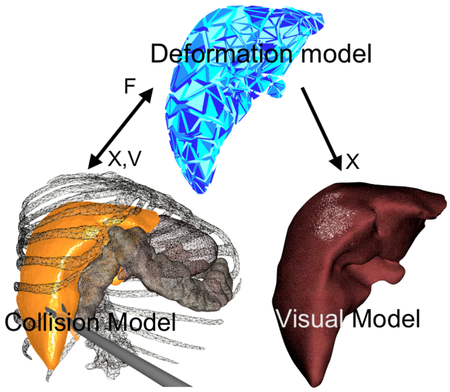


Figure 11. A simulated Liver. Three representations are used: one master model for the internal deformable mechanics, one for the collisions, and one for the visualization. Mappings (black arrows) are used to propagate positions (X) and velocities (V) from master to slaves, while forces (F) are propagated in the opposite direction

We continue the development of SOFA, the open source simulation library, which is becoming an international reference, and we have published a chapter on it in a Springer book [28]. SOFA facilitates collaborations between specialists from various domains, by decomposing complex simulators into components designed independently and organized in a scenegraph data structure. Each component encapsulates one of the aspects of a simulation, such as the degrees of freedom, the forces and constraints, the differential equations, the main loop algorithms, the linear solvers, the collision detection algorithms or the interaction devices. The simulated objects can be represented using several models, each of them optimized for a different task such as the computation of internal forces, collision detection, haptics or visual display, as illustrated in Figure 11 . These models are synchronized during the simulation using a mapping mechanism. CPU and GPU implementations can be transparently combined to exploit the computational power of modern hardware architectures. Thanks to this flexible yet efficient architecture, SOFA can be used as a test-bed to compare models and algorithms, or as a basis for the development of complex, high-performance simulators

6.3.7. Real time fluid animation on GPU

Participant: Martin Guay.

In collaboration with Manuel Vennier (Maverick, Inria), we developped a simple and fast method to animate fluids on the GPU. Inspired from the classical SPH method (Smooth Particles Hydrodynamics), we express a weekly compressible formulation for the fluid animation. Contrary to standard approaches, we fully developed the formulation on a grid, leading to an efficient GPU implementation. The method replace the implicit formulation of pressure by an explicit one based on density invariance. We therefore propose a method to simulate 3D Eulerian gaseous fluids in a single pass on the GPU. The results published in [22] are less accurate than a standard fluid simulation approach, but lead to real-time fluid-looking models (see fig. 12) which are practicable for video games or other interactive applications.

Figure 12. Example of fluid results obtained by our approach in [22].

6.4. Knowledge-based models for narrative design

Scientist in charge: Rémi Ronfard

Other permanent researchers: François Faure, Jean-Claude Léon, Olivier Palombi

6.4.1. Computational model of film editing

Participants: Remi Ronfard, Quentin Galvane.

Collaboration with the Mimetic team (Marc Christie) is continuing on this topic as part of the CINECITA (ANR jeune chercheur) and CHROME (ANR) projects.

We presented a survey of automatic video editing and new results from our ongoing collaboration at the first workshop on intelligent cinematography and editing (WICED) which took place during the Foundation of Digital Games (FDG) international conference [18], [14].

6.4.2. Stochastic Plex Grammars

Participant: Remi Ronfard.

During Quentin Doussot's master thesis, we experimented with stochastic plex grammars, which proved to be efficient for generating 3D scenes in the style of Keith Haring [17]. The model is able to generate static scenes by assembling colorful body parts into Keith Haring figures. The model is also able to simulate Markov chains of such figures by randomly changing attributes and composition of the scene.

6.4.3. Reframing theatre performances

Participants: Remi Ronfard, Vineet Gandhi.

In 2012, we made full-hd video recordings of rehearsals and performances at Celestins - Theatre de Lyon:

- A l'Ouest, directed by Nathalie Fillion. Coproduction - Théâtre du Rond-Point, Célestins, Théâtre de Lyon, Cie Théâtre du Baldaquin, AskUs , Le Gallia Théâtre-Saintes. Coproduction - Théâtre du Rond-Point, Célestins, Théâtre de Lyon, Cie Théâtre du Baldaquin, AskUs , Le Gallia Théâtre, Saintes.
- Lorenzaccio, directed by Claudia Stavisky, Théâtre de Saint Petersbourg.
- Mort d'un commis, directed by Claudia Stavisky, Célestins, Théâtre de Lyon.

As part of his PhD thesis, Vineet Gandhi developped novel algorithms for actor detection and naming. This has been tested on movies as well as theatre performances. Current work is focusing on automatically reframing those recordings into cinematically-valid shots focusing on one or more actors.

A related thread of work was started for semantic annotation of the recordings using the syntax and semantics of blocking notations, a symbolic notation used in North-American theatres [25].

6.4.4. Virtual theatre rehearsals

Participant: Remi Ronfard.

We are starting to investigate the possibility of rehearsing theatre plays with real and virtual actors, using extensions of interactive scores initially proposed for computer music. A position paper was presented to researchers in theatre studies during a seminar on the notation of theatre [24].

6.4.5. Extracting functional information from assembly models

Participants: Jean-Claude Léon, Ahmad Shahwan, Olivier Palombi.

Assembly models of products, as available from CAD software reduce to a set of independent geometric models of its components and a logical structure of the assembly described as a tree containing components' names. Such a model lacks of geometric connections between its components and the work performed at 6.5 contributes already to structure the geometric model of each component with its geometric interfaces. However, the assembly tree structure and components' names still have no connection with the geometric model of components and their names don't convey robust information because their are user chosen. Here, the purpose is to set tight connections between components' geometric models and their functions. Using dualities between geometric interfaces and interaction forces, it is possible to initialize qualitative mechanical values at each geometric interface, producing different possible configurations.

Then, the proposed approach builds upon relationships between function, behavior and shape to derive functional information from the geometry of component interfaces. Among these concepts, the concept of behavior is more difficult to set up and connect to the geometry of interfaces and functions. Indeed, states and design rules are introduced to express the behavior of components through a qualitative reasoning process [7]. This reasoning process, in turn, takes advantage of domain knowledge rules and facts, checking the validity of certain hypotheses that must hold true all along a specific state of the product's lifecycle, such as operational, stand-by or relaxed states. Eliminating configurations at geometric interfaces that contradict one or more of those hypotheses in their corresponding reference state reduces ambiguity, subsequently producing functional information in a bottom-up manner.
This bottom-up process starts with the generation of a Conventional Interfaces Graph (CIG) with components as nodes, and conventional interfaces (CI: the geometric interfaces) as arcs. A CI is initially defined by a geometric interaction that can be a contact or an interference between two components. CIs are then populated with Functional Interpretations (FI) according to their geometric properties, producing potentially many combinations. A first step of the reasoning process, the validation against reference states, reduces the number of FIs per CI. Then, a matching process takes place using inferences of an ontology reasoner to produce a functional designation of each component. The ontology is based on several taxonomies: conventional interfaces, functional interfaces and functional designations that are connected through the qualitative reasoning process. As a result, the geometric model of each component assigned with a functional designation becomes intrinsically structured with functional interfaces (see Figure 13). Structured models can

be used to perform high level shape transformations like virtual prototypes. MyCorporisFabrica is a software framework we started to connect to. This activity is part of the ANR ROMMA project. It is a first contribution to simulation scenarios.

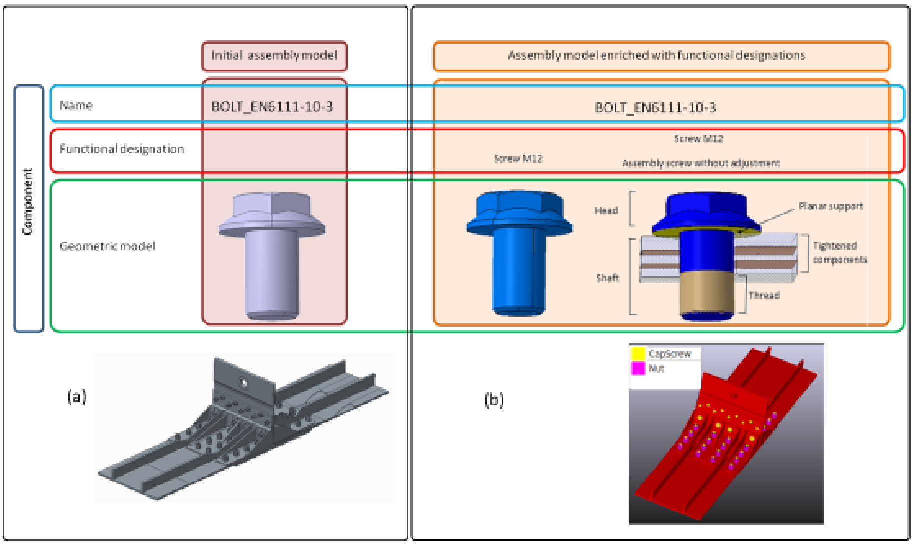


Figure 13. An example of assembly before (a) and after (b) the extraction of functional information. The upper part shows the influence of the extraction process on the structure of components' geometric models. The lower part illustrates the extraction process applied to a mechanical assembly.

6.4.6. Anatomical models

Participants: Ali-Hamadi Dicko, Olivier Palombi, François Faure.

We continue the development and the exploitation of MyCF, our ontology-centered anatomical knowledge base, in collaboration with the Grenoble University Hospital, and the DEMAR team in Montpellier (Benjamin Gilles).

We have presented a novel pipeline for the construction of biomechanical simulations by combining generic anatomical knowledge with specific data [27], [21], as illustrated in Figure 14 . Based on functional descriptors supplied by the user, the list of the involved anatomical entities (currently bones and muscles) is generated using formal knowledge stored in ontologies, as well as a physical model based on reference geometry and mechanical parameters. This simulation-ready model can then be registered to subject-specific geometry to perform customized simulations. The user can provide additional specific geometry, such as a simulation mesh, to assemble with the reference geometry. Subject-specific information can also be used to individualize each functional model. The model can then be visualized and animated. This pipeline dramatically eases the creation of biomechanical models.

6.4.7. Managing morphological and functional information of the human body

Participants: Olivier Palombi, Ali-Hamadi Dicko, François Faure, Jean-Claude Léon, Ahmad Shahwan.

My Corporis Fabrica (MyCF) is an anatomical knowledge database. During 2012, we have linked functional entities defined in MyCF to the involved anatomical structures. The scope has been limited to the musculoskeletal system. Based on this brain new formal description of the functional anatomy of limbs, we present a novel pipeline for the construction of biomechanical simulations by combining generic anatomical knowledge with specific data. This pipeline dramatically eases the creation of biomechanical models [27].

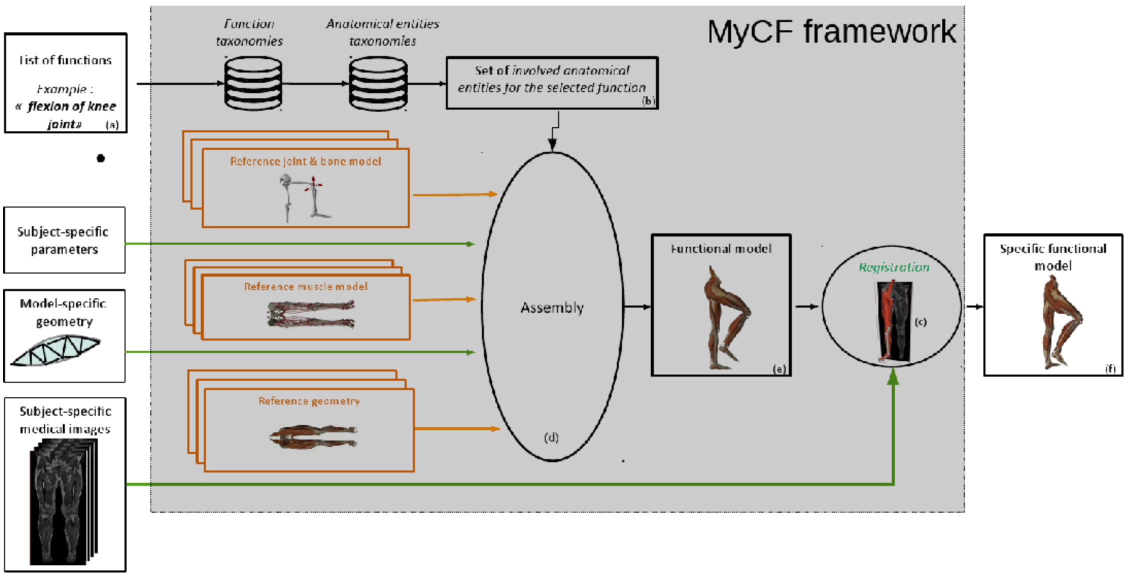


Figure 14. An overview of our modeling framework. On the left, the user input is a list of functions to simulate, optionally complemented with specific data. On the right, the output is a mechanical model ready for simulation. The modeling pipeline uses symbolic knowledge to select anatomical entities to assemble. The final model can be composed of a mix of reference and specific parameters and geometry.

MyCF-Browser which is the GUI of MyCF has been completely reviewed and rewritten. The MyCf's style software architecture is REST (Representational State Transfer) that has emerged as a predominant Web service design model. The anatomical knowledge is now available through a WEB service. The next step is to write a full web MyCF-Browser. MyCF browser is now available on line: http://www.mycorporisfabrica.org/. The MyCf's generic programming framework can be used for other domains. The link with semantic and 3D models matches research activities of IMAGINE towards interactive digital creation media. Anatomy can be seen as a study case.

6.5. Creating and interacting with virtual prototypes

Scientist in charge: Jean-Claude Léon

Other permanent researchers: Marie-Paule Cani, Rémi Ronfard, Olivier Palombi

6.5.1. Space deformations

Participant: Stefanie Hahmann.

Free Form Deformation (FFD) is a well-established technique for deforming arbitrary object shapes in space. Although more recent deformation techniques have been introduced, amongst them skeleton-based deformation and cage based deformation, the simple and versatile nature of FFD is a strong advantage, and justifies its presence in nowadays leading commercial geometric modeling and animation software systems. Several authors have addressed the problem of volume preserving FFD. These previous approaches however make either use of expensive non-linear optimization techniques, or resort to first order approximation suitable only for small-scale deformations. Our approach was to take advantage from the multi-linear nature of the volume constraint in order to derive a simple, exact and explicit solution to the problem of volume preserving FFD. Two variants of the algorithm have been developed, without and with direct shape manipulation.

Moreover, we showed that the linearity of our solution enables to implement it efficiently on GPU. This work has been done in collaboration with Gershon Elber from TECHNION, Hans Hagen from TU Kaiserslautern, Georges-Pierre Bonneau and Sébastien Barbier from Maverick Inria. It has been published in the journal The Visual Computer [5].

Figure 15. Comparison between standard FFD deformation (middle) and our method preserving the volume (right) from an initial rest shape (left).

Within Lucian Stanculescu PhD, we developped a mesh structure that dynamically adapts to the deformation defined by the user. Thanks to the quasi-uniform property of the mesh, it can be locally extended by any arbitrary deformation, and the mesh can also hanlde changes of topologies to be used as a virtual sculpting tool. This year we extend this work to handle local features such as sharp edges. In defining features (points or curves) over the surface we can interactively define meaningfull regions limiting the influence of the deformation tools, or to ease artistic decorration mapping such as textures or extra geometric layers. We aim to generate a new tool enabling to sculpt objects which blend between organic to CAD-style appearance.

6.5.2. Procedural modeling of terrains and cities

Participants: Adrien Bernhard, Marie-Paule Cani, Arnaud Emilien.

Within the PhD of Adrien Bernhard we introduced a real-time terrain modeling tool using a fast GPU-based terrain solver with a lightweight CPU-based data structure.

We then work on adding roads and settlements on this terrain within the PhD of Arnaud Emilien. We focused on the modeling of small, European villages that took benefit of terrain features to settle in safe, sunny or simply convenient places. We introduced a three step procedural method [3] for generating scattered settlements on arbitrary terrains, enabling villages and hamlets, with the associated roads, forests and fields to be built on arbitrary landscapes.

Figure 16. Fortified village at the top of a cliff, using a war-time growth scenario followed by farming style settlement.

6.5.3. Hand Navigator

Participant: Jean-Claude Léon.

The different deformation models we developed in the past few years open the problem of providing intuitive interaction tools for specifying the desired deformations in real-time. Therefore, work has focused on developing new devices to investigate interactions incorporating a rather large number of parameters. For the past three years, we focused on developing a peripheral device similar to a mouse, called the HandNavigator, enabling to control simultaneously ten or more degrees of freedom of a virtual hand. This device developed in collaboration with Jean-Rémy Chardonnet (Inst. Image, Arts et Métiers ParisTech) consists in a 3D mouse for the position and orientation of the hand in 3D space, enhanced with many sensors for moving and monitoring the virtual fingers. Thanks to a pre-industrialization project funded by the incubator GRAVIT, the first prototype, patented by Inria, has been extended with the incorporation of new sensors and new shapes to improve the device efficiency and evolve toward a passive haptic device (see Figure 17). An extension of the patent and a partnership with HAPTION company are new steps toward the industrialization of this device. The partnership with HAPTION focuses on grasping actions to use the Hand Navigator as a complement to their haptic feedback device. Publications took place after setting up the patent extension [12], [13]. The ongoing BQR INTUACTIVE funded by Grenoble-INP will lead to further scientific topics regarding interactions during grasping as well as with deformable bodies and a partnership is ongoing with GIPSA-Lab to study the muscular activity during interactions. A specific experiment has been set up to study the user's muscles activity.

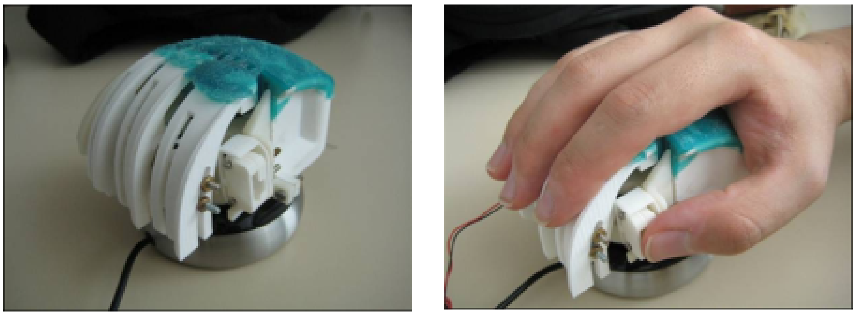


Figure 17. Current version of the HandNavigator prototype with three sensors per finger and a vibration damping structure.

IMARA Project-Team

6. New Results

6.1. Low speed automation

Participants: Paulo Lopes Resende, Fawzi Nashashibi, Hao Li, Evangeline Pollard.

The ABV project builds on the HAVEit philosophy (a previous IMARA project for high speed automation) of offering higher levels of automation on highways and organizing the cooperation between human and system along novel automation levels. It differs from HAVEit by focusing on congested traffic at speeds below 50 km/h and adding fully automated driving to the automation spectrum. By automatically following congested traffic, the ABV system relieves the human driver from monotonous tasks. During fully automated driving, the human driver is not required to monitor the system, but has to take over control at the end of the application zone.

6.2. Urban autonomous driving: dealing with intersections

Participants: Guillaume Tréhard, Evangeline Pollard, Fawzi Nashashibi.

The goal of this project, made in collaboration with Valeo is to develop a complete solution for autonomous driving on open roads. More specifically, IMARA's objectives are to provide the way to safely cross any kind of intersections for an autonomous vehicle in a urban context. Among the different relevant scenarios, we can notice:

- Intersection with different shapes: Roundabout, T junctions , X junctions;
- Intersection with different rules: With specific rules (traffic lights, main road...) or unspecified rules ("priority to the right");
- Different traffic: Busy or empty intersections;
- Deal with abnormal situations: road works, policemen, firemen,...

Possible steps for this work can be listed as follows:

- Model the intersection: define relevant information, find a generic model for every intersection;
- Detect the intersection (shape, drivable area, traffic flows);
- Understand the priorities that rules it;
- Locate the car in the intersection by crossing it;
- Plan a path to get out of the intersection.

6.3. Conception of a new communicative system for the protection of vulnerable people

Participants: Pierre Merdrignac, Evangeline Pollard, Oyunchimeg Shagdar, Fawzi Nashashibi.

A new research has been recently launched at IMARA team. The goal is to elaborate a new communicative system between vulnerable people (pedestrian, person with reduced mobility, bicyclist, etc.) and intelligent vehicles in order to improve safety and to limit collision risk. The main idea of this project is as follows. Intelligent vehicles are equipped with an obstacle detection/classification /tracking module in order to prevent injuries. On the other hand, to help the driver in this challenging task, vulnerable people use an application on their mobile phone to inform/share their status on location, type, and dynamics. The status information is transmitted to the driver utilizing wireless communications technology (e.g., 3G and Wi-Fi). In the vehicle, information coming from the communications device and obstacle detection module will be merged to improve the detection and classification tasks. In case of emergency, the vehicle can broadcast safety information to vulnerable people.

6.4. Visible Light V2V Communications for Platooning Control

Participants: Mohammad Abu Alhoul, Oyunchimeg Shagdar, Mohamed Marouf, Fawzi Nashashibi.

Fully automated vehicles have the potential to greatly improve the comfort of humans' life. For driving from one place to another, an automated vehicle must avoid collisions and be able to select non-congested roads for safe and efficient driving. In order to do that the vehicle needs to control its mobility in both macroscopic and microscopic levels by utilizing information exchange with other vehicles and roadside infrastructures based on wireless communications technology. While radio frequency channel is a convincing choose for vehicular communications due to its high data rate over relatively long coverage range (minimum several 100's meters), it is expected to experience channel congestion and low communication reliability especially for the scenario where there is high-density vehicles. In such scenarios vehicles still require to control the mobility on both the macroscopic and microscopic levels, we need to look for supportive and at the same time practical communication media with the ability to support sufficient connection between vehicles. According to the latest standard from IEEE, 802.15.7 for 2011, the communication coverage of Visible Light Communication (VLC) can reach up to 10's of meters, which match the information exchange requirements for mobility control in microscopic level. Motivated by this we started our research activity on modeling of visible light communications channel and design of microscopic mobility control, specifically platooning control, using VLC.

6.5. Augmented reality for the protection of vulnerable people

Participants: Hao Li, Fawzi Nashashibi.

A brand new idea of cooperative augmented reality is under development in IMARA team. It utilizes the results of cooperative local mapping to realize certain augmented reality effect. More specifically, the idea is to obtain an augmented effect of "seeing" through front vehicle, based on the intelligent vehicle sensor configurations.

Given a scenario of two vehicles: a front (first) vehicle and a following (second) vehicle. This front-following vehicles scenario is typical in traffic environment and is potentially dangerous, especially in some occasions such as during an overtaking, where the front vehicle occludes a part of the scene to the following vehicle. The idea of cooperative augmented reality is thus to project the visual perception of the front vehicle onto that of the following vehicle, abiding by perspective geometry. In other words, we patch the occluded part of the view of the following vehicle with corresponding part of the view of the front vehicle. This is not simply a process of partial view copying and pasting between the two vehicles; we have to transform the partial view of the front vehicle according to perspective geometry, in order to make a vivid and natural reproduction of this partial view for the following vehicle, just like if the following vehicle can directly see into the occluded area.

A prerequisite for performing the perspective transformation between the visual perceptions of the two vehicles is the knowledge of the visual perception depth. This knowledge can be estimated by stereo-vision, if correct correspondence is established (yet a challenging process) between the images pair in stereo-vision. However, approximate estimate of the visual perception depth was obtained with the help of 2D range perception in an innovative way and indirect vehicle-to-vehicle relative pose estimation method introduced in [36].

6.6. Step detection for Personal Mobility Vehicles

Participants: Evangeline Pollard, Joshué Pérez Rastelli, Fawzi Nashashibi.

Personal Mobility Vehicles (PMV) is an important part of the Intelligent Transportation System (ITS) domain. These new transport systems have been designed for urban traffic areas, pedestrian streets, green zones and private parks. In these areas, steps and curbs make the movement of disable or mobility reduced people with PMV, and with standard chair wheels difficult. In this work, a step and curb detection system based on laser sensors has been developed. This system is dedicated to vehicles able to cross over steps, for transportation systems, as well as for mobile robots. The system is composed of three laser range finders. Hokuyo UTM 30 LX devices were chosen for their large detection angle (270°) and their high angular resolution (0.25°) and range (30m).

Two laser sensors dedicated to the step detection have a vertical orientation in order to scan the altitude profile of the environment over two lines of sight and the third one, with a vertical orientation is dedicated to obstacle detection.

The step detection process is thus based on the study of the first derivative of the altitude and highlights the use of a new algebraic derivative method (Alien) adapted to laser sensor data. The system has been tested on several real scenarios. It provides the distance, altitude and orientation of the steps in front of the vehicle and offers a high level of precision, even with small steps.

6.7. PROSIVIC-CTS simulator

Participant: Joshué Pérez Rastelli.

The Architecture validation and experiments presented in this document have been implemented in a simulated environment, called ProSiVIC, which allows implementing a virtual Cybercars, among other vehicles. The algorithms are the same as in our Cybus platform, using the position from the SLAM and DGPS sensors. The ProSivic simulator offers a multi-sensorial environment, and takes into account several parameters of a real car such as the inertia, steering wheel response, lateral acceleration with yaw angles, damping suspension, simple weather conditions, friction parameters and more.

Moreover, synchronized time, acceleration (in wheel torque), steering, odometer information, lidar information and camera viewports are some of the components supporting the connection between the control architecture in RTMaps and the simulation.

The simulations show the behavior of the control architecture implemented for CTSs. Two urban scenarios were tested: roundabouts and intersections.

6.8. Autonomous docking based on infrared system for electric vehicle charging in urban areas

Participants: Benjamin Lefaudeux, Joshué Pérez Rastelli, Fawzi Nashashibi.

One of the recent aims of the Intelligent Transportation Systems (ITS) is the reduction of air pollution, reducing the fuel consumption in urban areas and improving road security. To this purpose, electric vehicles are a good and high demanded alternative. Nowadays, some big cities are launching the first electric car-sharing projects to clear its traffic jam, as an alternative to the classic public transportation systems. However, there are still some problems related to energy storage, charging and autonomy to be solved. To tackle this problem in the context of the French project AMARE, IMARA has developed an autonomous docking system, based on an infrared camera embarked in a vehicle equipped with dedicated ADAS, and some infrared diodes installed in the infrastructure, for recharging the vehicle batteries in a street parking area. The results obtained show a good behavior of the implemented system, which is working in a real scenario in the city of Paris.

Different experiments, departing from different points, show a good behavior of the proposed systems. Both lateral and longitudinal errors are lower that the limits of the charging station. The controller used is easy and intuitive for tuning, and the gains can be adapted according of the different vehicles characteristics. This technology permits to assist to human drivers in the charging process of electric vehicles in cities.

6.9. Reasoning for relaxing traffic regulations

Participants: Philippe Morignot, Fawzi Nashashibi.

This work [39] deals with relaxation of traffic rules in unusual but practical situations. For example, if a truck is unloading on a roadway, the automated vehicle should overtake it despite a continuous yellow line: traffic rules are indeed broken, which is illegal, but this might be tolerated due to the unusual aspect of the situation at hand.

An ontology has been developed in order to represent the road network (a directed graph, vertices being intersections and edges being lanes), the infrastructure (road signs, marks), the other road users and the intelligent vehicle. Reasoning on this representation is performed by inference rules (IF/THEN symbolic structures), encoding the deliberation on the encoded situation. Main rules conclude on the next discrete motion of the vehicle, e.g., “pass onto the adjacent lane” which involves crossing a continuous yellow line.

In practice, this ontology has been created using the PROTEGE ontology editor from Stanford University. IF/THEN rules are represented in SWRL (Semantic Web Rule Language), using the reasoner PELLET from the company Clark & Parsia (a plug-in of the tool PROTEGE).

Work over the next year involves porting this reasoning module on the vehicles: porting the generated JAVA source code as one component inside the RTMAPS architecture of CyberCars, and linking the ABoxes (assertional boxes) of the ontology to symbols extracted from signals by perception.

6.10. Communications and Management Control for Cooperative Vehicular Systems

Participants: Ines Ben Jemaa, Oyunchimeg Shagdar, Fawzi Nashashibi, Arnaud de La Fortelle.

One of the attractive applications of electric autonomous vehicles is electric automated Car Sharing service. In this application, a user requests a vehicle at a given geographical location triggering the car sharing system to allocate an autonomous vehicle for the user transport from the station to the user’s desired destination. The application requires efficient cooperation among the autonomous vehicles and a service management centre for reliable and responsive car sharing service. Such cooperation is not possible unless vehicles exchange their information on e.g., position, motion, and coordination messages among themselves and with central management entities. While the existing wireless communications technologies can be applied for vehicle to vehicle and vehicle to the infrastructure communications, important research challenges remain including network partitioning problem caused by vehicles’ mobility and inability of the convergence of geographically scoped V2V and Internet-based V2I communications. Targeting these issues, we study a topology control solution to tackle the network portioning problem and design of unicast/multicast/Geonetworking schemes for convergence of V2V and V2I communications systems for car-sharing applications [28].

6.11. New urban transportation platforms: Inria’s Cybus

Participants: Laurent Bouraoui, François Charlot, Fawzi Nashashibi, Paulo Lopes Resende, Michel Parent, Armand Yvet.

Cybus is the newest prototyping and demonstration platform designed at Inria. Apart from the chassis and engines, the whole hardware and software systems were developed thanks to IMARA’s researchers and engineers talents. These electric vehicles are based on a Yamaha chassis but the embedded intelligence is the result of two years of development. Much of the perception and control software has been improved. New guidance functionalities were developed this year, mainly with the introduction of stereovision-based SLAM.

The platforms developed here (Cybus) will be demonstrated in the context of the EU CityMobil-2 project. This time real operational mobility services demonstrations will be extended to 6-12 months in selected European cities! Other showcases are expected to take place in Asian cities in 2014.

6.12. Belief propagation inference for traffic prediction

Participants: Cyril Furtlehner, Yufei Han, Jean-Marc Lasgouttes, Victorin Martin.

This work [57] deals with real-time prediction of traffic conditions in a setting where the only available information is floating car data (FCD) sent by probe vehicles. The main focus is on finding a good way to encode some coarse information (typically whether traffic on a segment is fluid or congested), and to decode it in the form of real-time traffic reconstruction and prediction. Our approach relies in particular on the belief propagation algorithm.

Figure 1. The Cybus operated at La Rochelle City during 3 months as a free transport service.

These studies are done in particular in the framework of the projects Travesti and Pumas.

This year's highlights are

- A paper describing a new sufficient condition for local stability of the Belief Propagation algorithm has been published and presented in an international conference [38].
- The work about the theoretical aspects of encoding real valued variables into a binary Ising model has been summarized in a publication currently under reviewing process.
- Ideas about adding macroscopic variables within the Ising model are currently being tested using the software BPstruction developed last year.
- Victorin Martin has given a talk at the "Colloque Jeunes Probabilistes et Statisticiens" at CIRM, where he presented his work one the design of a latent Ising model for real valued inference.
- Cyril Furtlehner, Yufei Han and Victorin Martin presented the work done in the Travesti project at the workshop on inference organized by Inria and Mines ParisTech (see 9.1).

6.13. Non-negative Tensor factorization for spatio-temporal data analysis

Participant: Yufei Han.

This is a joint work with Fabien Moutarde from Mines ParisTech.

We investigate the use of non-negative tensor factorization for spatio-temporal data clustering and prediction. In general case, a spatio-temporal signal is represented as a set of multiple-variant temporal sequences. In the domain of intelligent traffic, the temporal records of traffic flow states (free-flowing/congestion) over a specific time duration with respect to hundreds of links in a transportation network can be considered as a simple but direct example of spatio-temporal signal. Both temporal causality between neighboring time sampling steps and spatial layout of the multiple-variant observation captured at each time sampling step are the focus of the spatio-temporal data analysis. Non-negative tensor factorization enables us to project the high dimensional spatio-temporal data into low-dimensional subspace and clustering/prediction can be then achieved on the derived subspace projection easily.

This year's highlights are

- A conference paper describing application of non-negative tensor factorization in traffic flow state prediction and clustering has been published and presented at ITS World Congress [30];
- The application of non-negative matrix factorization in clustering network-level traffic flow state in large-scale transportation network has been accepted for publication in a journal [11].

6.14. Sparse covariance inverse estimate for Gaussian Markov Random Field

Participants: Cyril Furtlehner, Yufei Han, Jean-Marc Lasgouttes, Victorin Martin.

We investigate in [53] different ways of generating approximate solutions to the inverse problem of pairwise Markov random field (MRF) model learning. We focus mainly on the inverse Ising problem, but discuss also the somewhat related inverse Gaussian problem. In both cases, the belief propagation algorithm can be used in closed form to perform inference tasks. We propose a novel and efficient iterative proportional scaling (IPS) based graph edit method to identify sparse graph linkage of GMRF model to fit underlined data distribution. We remark indeed that both the natural gradient and the best link to be added to a maximum spanning tree solution can be computed analytically. These observations open the way to many possible algorithms, able to find approximate sparse solutions compatible with belief propagation inference procedures and sufficiently flexible to incorporate various spectral constraints like e.g. walk summability. Experimental tests on various data sets with refined L_0 or L_1 regularization procedures indicate that this approach may be a competitive and useful alternative to existing ones.

The part of this work dedicated to Gaussian Markov Random Field has been submitted to the AISTATS 2013 conference.

6.15. Evaluation of dual mode transport system by event-driven simulation

Participants: Arnaud de La Fortelle, Jean-Marc Lasgouttes, Thomas Liennard.

The European project CATS — City Alternative Transport System — is developing and evaluating a new vehicle system using a single type of vehicle for two different usages: individual use or collective transport. Real experiments will necessarily take place with a limited number of vehicles and stations. Hence there is a need for evaluation using simulations. We have been developing a discrete events simulator for that purpose, based on a previous work done for collective taxis [58].

Our model relies on an adapted events/decision graph that extends previous graphs. The new feature of this model is the way we deal with two modes that can be extended to many other modes. This work therefore shows on a concrete example a method to efficiently merge multiple modes into one model.

This year has seen the overhaul of the simulator implementation, as well as the development of a result visualizer that can replay the simulations on a map and show various statistics.

6.16. Multi-speed exclusion processes

Participants: Cyril Furtlehner, Jean-Marc Lasgouttes.

The slow-to-start mechanism is known to play an important role in the particular shape of the fundamental diagram of traffic and to be associated to hysteresis effects of traffic flow. We study this question in the context of stochastic processes, namely exclusion and queueing processes, by including explicitly an asymmetry between deceleration and acceleration in their formulation. Spatial condensation phenomena and metastability are observed, depending on the level of the aforementioned asymmetry. The relationship between these 2 families of models is analyzed on the ring geometry, to yield a large deviation formulation of the fundamental diagram (FD)

This work has been published in the Journal of Statistical Physics [10].

6.17. Herding behavior in a social game

Participants: Guy Fayolle, Jean-Marc Lasgouttes.

The system *Ma Micro Planète* belongs to the so-called *Massively Multi-Player online Role Playing game* (MMORPG), its main goal being to incite users to have a sustainable mobility. Two objectives have been pursued.

- Construct an experimental platform to collect data in order to prompt actors of the mobility to share information (open data system).
- See how various mechanisms of a game having an additive effect could modify the transportation requests.

At the heart of the game are community-driven *points of interest* (POIs), or *sites*, which have a score that depends on the players activity. The aim of this work is to understand the dynamics of the underlying stochastic process. We analyze in detail the stationary regime of the system in the thermodynamic limit, when the number of players tends to infinity. In particular, for some classes of input sequences and selection policies, we provide necessary and sufficient conditions for the existence of a complete meanfield-like measure, showing off an interesting *condensation* phenomenon.

The work has been completed during this year [51] and has been submitted to a journal for publication.

6.18. Exact asymptotics of random walks in the quarter plane

Participant: Guy Fayolle.

In collaboration with K. Raschel (CNRS, Université F. Rabelais à Tours), we pursued the works initiated in 2011.

The enumeration of planar lattice walks, is a classical topic in combinatorics. For a given set $\mathcal{S}$ of allowed unit jumps (or steps), it is a matter of *counting the number of paths* starting from some point and ending at some arbitrary point in a given time, and possibly restricted to some regions of the plane.

Like in the probabilistic context, a common way of attacking these problems relies on the following analytic approach. Let $f(i,j,k)$ denote the number of paths in $\mathbb{Z}_+^2$ starting from $(0,0)$ and ending at (i,j) at time k. Then the corresponding CGF

$$F(x,y,z) = \sum_{i,j,k \geq 0} f(i,j,k) x^i y^j z^k$$

satisfies the functional equation

$$K(x,y)F(x,y,z) = c(x)F(x,0,z) + \widetilde{c}(y)F(0,y,z) + c_0(x,y),$$

where x,y,z are complex variables, although the time variable z plays somehow the role of a parameter. The question of the type of the associated counting generating functions, that is rational, algebraic, holonomic (solution of a linear differential equation with polynomial coefficients), was solved whenever the group is *finite* (see RA 2010). When the group is infinite, the problem is still largely.

It turns out that the nature of singularities play a deep important role in this classification. Making use of the general and powerful approach proposed in the book [2], the paper [9] has been presented at the 23rd International Conference *AofA 2012* on *Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, Montreal, June 17-22.

6.19. Statistical physics and hydrodynamic limits

Participant: Guy Fayolle.

These last years, having in mind a global project concerning the analysis of complex systems, we did focus on the interplay between discrete and continuous description: in some cases, this recurrent question can be addressed quite rigorously via probabilistic methods (see previous activity reports).

To describe the systems of interest, which are in touch with many application domains, we started from *paradigmatic* elements, namely discrete curves subject to stochastic deformations. Up to some convenient mappings, it appears that most models can be set in terms of interacting exclusion processes, the ultimate goal being to derive *hydrodynamic limits* after proper scalings.

The key ideas can be found in [56], where the basic ASEP system on the torus is the toy model. In this case, the usual sequence of empirical measures, converges in probability to a deterministic measure, which is the unique weak solution of a Cauchy problem.

The Gordian knot is indeed the analysis of a family of specific partial differential operators in infinite dimension. Indeed, the values of functions at given points play here the role of usual variables, their number becoming infinite. The method presents some new theoretical features, involving path integrals, promeasures (as introduced by Bourbaki), variational calculus, and the construction of *generalized measures*. In [56], we present a detailed analysis of the ASEP system on the torus $\mathbb{Z}/N\mathbb{Z}$. Then we claim that most of the arguments a priori for multi-type exclusion processes, and should lead to systems of coupled partial differential equations of Burgers' type. At the moment, this claim is being proved for the famous ABC model, reformulated in terms of the dynamics of a random walk on the triangular lattice.

IMEDIA2 Team

6. New Results

6.1. Feature space modeling

Participants: Vera Bakic, Nozha Boujemaa, Esma Elghoul, Hervé Goëau, Sofiene Mouine, Olfa Mzoughi, Anne Verroust-Blondet, Itheri Yahiaoui.

6.1.1. Spatial relations between salient points on a leaf

Participants: Sofiene Mouine, Itheri Yahiaoui, Anne Verroust-Blondet.

We have introduced a novel method for leaf species identification combining local and shape-based features. Our approach extends the shape context model in two ways:
- First of all, two different sets of points are distinguished when computing the shape contexts: the voting set, i.e. the points used to describe the coarse arrangement of the shape and the computing set containing the points where the shape contexts are computed.
Three shape descriptors are proposed, as illustrated in Figure 1 : SC0 (spatial relations between margin points), which corresponds to the original shape context; SC1 (spatial relations between salient points) where the voting set and the computing set are composed of the salient points of the image and SC2 (spatial relations between salient and margin points) where the voting set contains the margin points and the computing set consists of the salient points (see [11] for more details).

Figure 1. From left to right: points used in SC0, SC1 and SC2. The small circles represent the sample points on the leaf margin. The cross points represent the salient points computed with Harris detector.

-This representation is enriched by introducing local features computed in the neighborhood of the computing points.
We obtained excellent identification scores in the ImageCLEF 2012 plant identification task for scan and scan-like images of leaves (RUN2 in [20]).

6.1.2. Detection and extraction of leaf parts for plant identification

Participants: Olfa Mzoughi, Itheri Yahiaoui, Nozha Boujemaa.

Automatic plant identification is a relatively new research area in computer vision that has increasingly attracted high interest as a promising solution for the development of many botanical industries and for the success of biodiversity conservation. Most of the approaches proposed are based on the analysis of morphological properties of leaves. They have applied several well-known generic shape descriptors. Nevertheless, faced with the large amount of leaf species, botanical knowledge, especially about leaf parts (petiole, blade, insertion point, base and apex, rachis) is important to enhance their precision.
First of all, in order to extract them from leaf images, we introduced two types of symmetry in [12]: (i) the local translational symmetry, which is useful for petiole and rachis detection and (ii) the local symmetry of depth indentations, which is suited for base and apex detection.
Then, we studied the usefulness of parts detection (mainly petiole and insertion point) as a pre-processing stage for classic leaf shape retrieval schemes [13]. We showed that the removal of the petiole and the use of the insertion point as a starting point for the descriptors sensitive to the starting point improve retrieval results.

6.1.3. Multi-organ plant identification

Participants: Hervé Goëau, Vera Bakic, Souheil Selmi.

Inspired by citizen sciences, the main goal of this work is to speed up the collection and integration of raw botanical observation data, while providing to potential users an easy and efficient access to this botanical knowledge. We therefore designed and developed an original crowd-sourcing web application dedicated to the access of botanical knowledge through automated identification of plant species by visual content with multi-organ queries. Technically, the first side of the application deals with multi-organ content-based identification of plant. Indeed, most methods proposed for such automatic identification are actually based on leaf images, mostly based on leaf segmentation and boundary shape. However, leaves are far from being the only discriminant visual key between species and they are not visible all over the year for a large fraction of plant species. We propose to make the use of five different organs and plant's views including habit, flowers, fruits, leaves and bark. Thanks to an interactive and visual query widget, the tagging process of the different organs and views is as simple as drag-and-drop operations and does not require any expertise in botany.

Technically, as suggested by the results of ImageCLEF2011 for leaves [24], it is based on local features and large-scale matching. Interest points are detected with a modified color Harris detector, in order to favor points with a central position in the image and to reduce the impact of background features. Each interest point is then described with a SURF local feature and an HSV histogram. Automatic system-oriented and human-centered evaluations of the application show that the results are already satisfactory and therefore very promising in the long term to identify a richer flora. The second side of the application deals with interactive tagging and allows any user to validate or correct the automatic determinations returned by the system. Overall, this collaborative system enables the automatic and continuous enrichment of the visual botanical knowledge and therefore it increases progressively the accuracy of the automated identification. This application called 'Identify' (cf. Figure 2 and http://identify.plantnet-project.org) has been presented at the first ACM International workshop on Multimedia Analysis and Ecological Data [8]. This work has been done in collaboration with Inria team ZENITH and with the botanists of the AMAP UMR team (CIRAD). It is also closely related to a citizen science project around plant's identification that we developed with the support of the Tela Botanica social network inside the Pl@ntNet project.

6.1.4. Segmentation transfer method for articulated models

Participants: Esma Elghoul, Anne Verroust-Blondet.

Mesh segmentation consists in partitioning the surface into a set of patches that are uniform with respect to a given property. We are interested in retaining the semantic information during the segmentation. A particularly challenging task is then the automatic identification of semantically meaningful parts of a 3D model, which can be hard to achieve when only the shape geometry is considered. We have introduced a method using a pre-segmented example model to perform semantic-oriented segmentation of non-rigid 3D models of the same class (human, octopus, quadrupeds, etc.). Using the fact that the same type of non-rigid models share the same global topological structure, we exploit coarse topological shape attributes in conjunction with a seed-based segmentation approach to transfer a meaningful and consistent segmentation from the example mesh to the

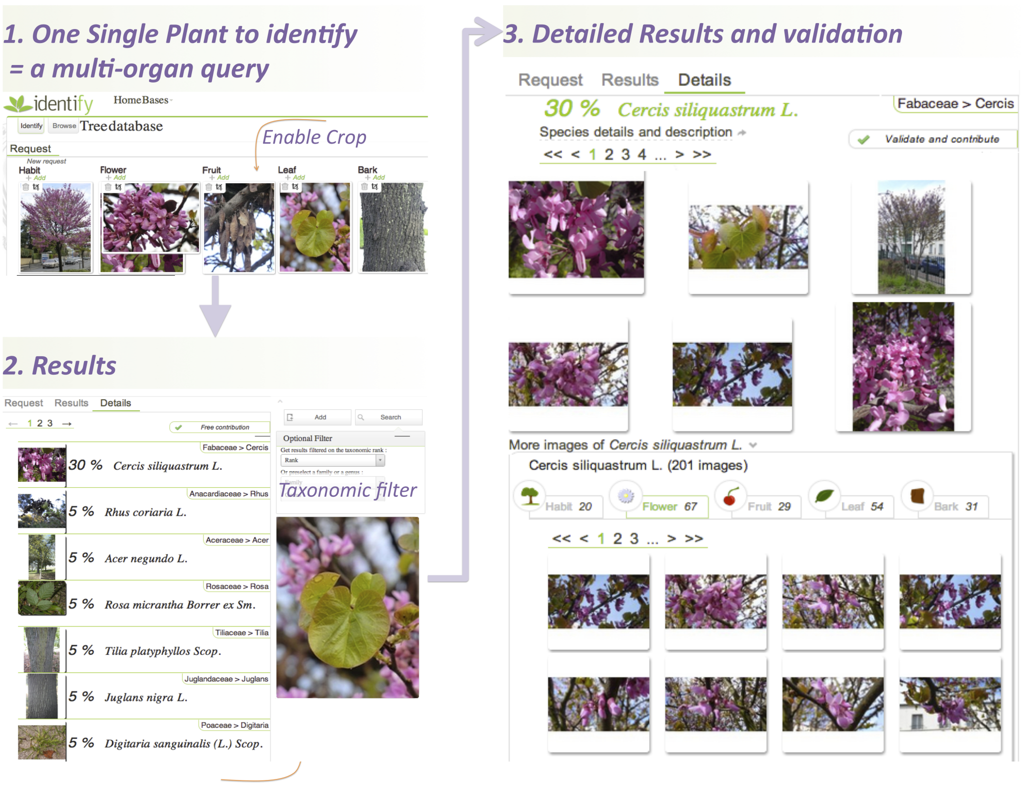


Figure 2. Example of a multi-organ query on one single plant submitted in the application.

target models. Promising results have been obtained on classes of articulated models (cf. Figure 3). This work has been submitted for publication.

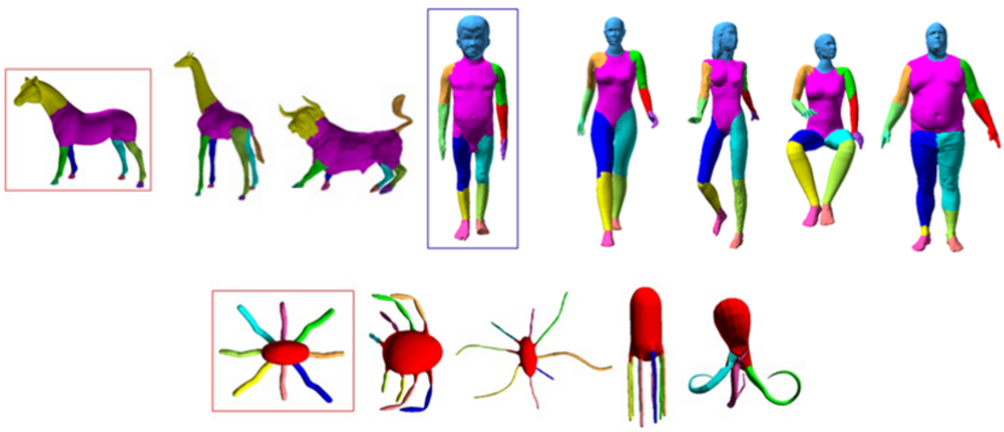


Figure 3. Segmentation transfer results on the quadrupeds, on the humans and on the octopus class. The exemplar segmentations are framed in red or in blue.

6.2. Feature space structuring

Participants: Nozha Boujemaa, Hervé Goëau, Amel Hamzaoui, Saloua Ouertani-Litayem, Mohamed Riadh Trad.

6.2.1. Plant Leaves Morphological Categorization with Shared Nearest Neighbors Clustering

Participants: Amel Hamzaoui, Hervé Goëau, Nozha Boujemaa.

In [9] we present an original experiment aimed at evaluating if state-of-the-art visual clustering techniques are able to automatically recover morphological classifications built by the botanists themselves. The clustering phase is based on a recent Shared-Nearest Neighbors (SNN) clustering algorithm, which allows combining effectively heterogeneous visual information sources at the category level. Each resulting cluster is associated with an optimal selection of visual similarities, allowing discovering diverse and meaningful morphological categories even if we use a blind set of visual sources as input. Experiments have been performed on ImageCLEF 2011 plant identification dataset [23], specifically enriched in this work with morphological attributes tags (annotated by expert botanists). The results presented in Figure 4 are very promising, since all clusters discovered automatically can be easily matched to one node of the morphological tree built by the botanists. This work is also described in details in Amel Hamzaoui's thesis [4].

6.2.2. Distributed KNN-Graph approximation via Hashing

Participants: Mohamed Riadh Trad, Nozha Boujemaa.

High dimensional data hashing is essential for scaling up and distributing data analysis applications involving feature-rich objects, such as text documents, images or multi-modal entities (scientific observations, events, etc.). In this first research track, we first investigated the use of high dimensional hashing methods for efficiently approximating K-NN Graphs [16], [19], [17], particularly in distributed environments. We highlighted the importance of balancing issues on the performance of such approaches and show why the baseline approach

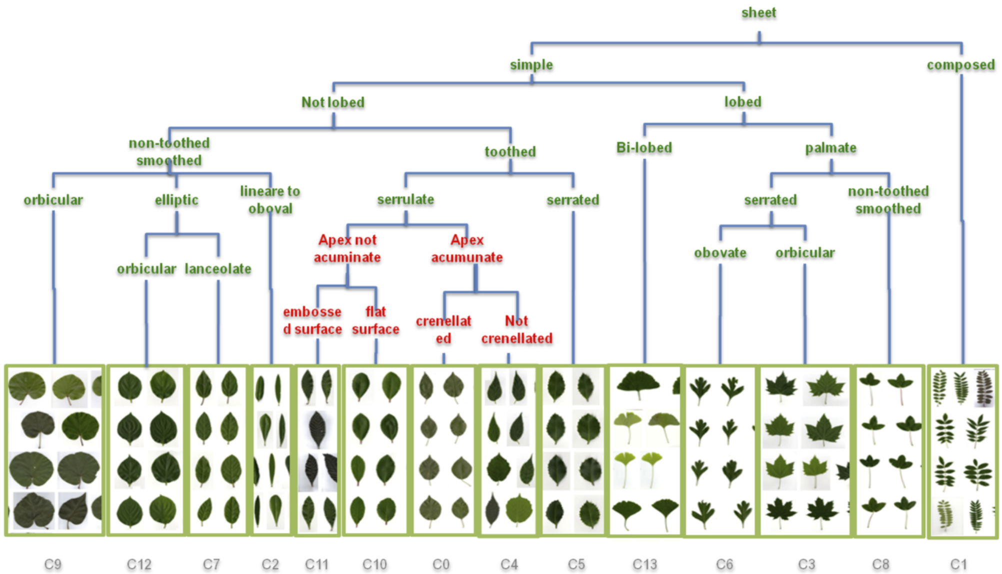


Figure 4. Hierarchical tree organization of the clusters produced by the SNN clustering method on the scan leaf dataset ImageCLEF2011.

using Locality Sensitive Hashing does not perform well. Our new KNN-join method is based on RMMH, a hash function family based on randomly trained classifiers that we introduced in 2011. We show that the resulting hash tables are much more balanced and that the number of resulting collisions can be greatly reduced without degrading quality. We further improve the load balancing of our distributed approach by designing a parallelized local join algorithm, implemented within the MapReduce framework.

6.2.3. Hash-Based Support Vector Machines Approximation for Large Scale Prediction

Participants: Saloua Ouertani-Litayem, Nozha Boujemaa.

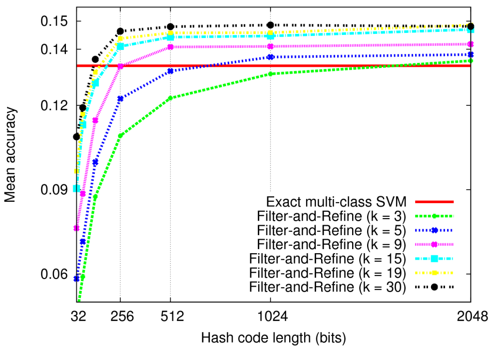
How-to train effective classifiers on huge amount of multimedia data is clearly a major challenge that is attracting more and more research works across several communities. Less efforts however are spent on the counterpart scalability issue: how to apply big trained models efficiently on huge non annotated media collections ? In [10], we addressed the problem of speeding-up the prediction phase of linear Support Vector Machines via Locality Sensitive Hashing. We proposed building efficient hash-based classifiers that are applied in a first stage in order to approximate the exact results and filter the hypothesis space. Experiments performed with millions of one-against-one classifiers show that the proposed hash-based classifier can be more than two orders of magnitude faster than the exact classifier with minor losses in quality (cf. Figure 5).

6.3. Pattern recognition and statistical learning

Participants: Nozha Boujemaa, Michel Crucianu, Donald Geman, Wajih Ouertani, Asma Rejeb Sfar.

6.3.1. Machine identification of biological shapes

Participants: Asma Rejeb Sfar, Donald Geman, Nozha Boujemaa.



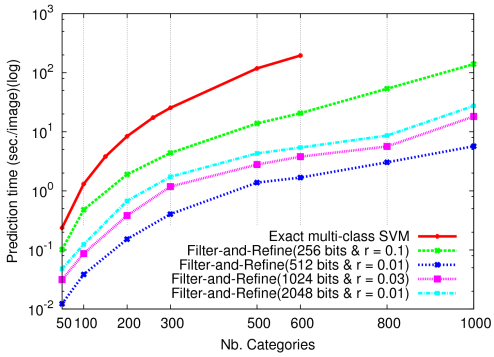


Figure 5. Exact Multi-class SVM vs HBMS based Filter-And-Refine method in terms of accuracy and prediction time

The increasing availability of digital images in the traditional sciences, the growing interest in biodiversity and the ongoing shortage of skilled taxonomists combine to make the automated categorization algorithms, increasingly important in many fields such as botany, agriculture and medicine. In this work, we propose a hierarchical coarse-to-fine approach to identify botanical species from a scanned sample of a plant organ, e.g., a leaf or a flower. To this end, we exploit domain-specific knowledge about taxonomy and landmarks. Promising recognition rates are achieved on several leaf datasets. Results have been submitted for publication.

6.3.2. Relevance feedback on partial image query

Participants: Wajih Ouertani, Michel Crucianu, Nozha Boujemaa.

scalability, hashing, SVM, prediction, approximation

Even if cropping an image to perform one-shot partial query filters a considerable amount of senseless regions for target definition, it does not yet clearly illustrates what the user is looking for. Indeed, the user target is either closer to the instance level or to the category level. Then we may have numerous suggested examples within the first response ranks while possibly some of them are totally irrelevant examples.

We claim that a localization interaction is still more appropriate than having a holistic decision about image relevance if it is performed on more examples. We go beyond the first partial query and investigate machine learning process to learn intention iteratively and interactively. Our learning process is based on what user delimit within additional images taken from the first response ranks. Our motivations include dealing with semantic gap revealed by local features hit falling into false regions within retrieved images. Those images might be either totally irrelevant, where all partial zones are out of the interest, or partially relevant, not because of the zones expected by the system (false-localization) but rather because of some missed zones. Through local annotations we expect the ability of redirecting the recognition session to those relevant regions and studying how much we can reduce the semantic gap within interactive localization.

This year, we studied several learning strategies based on several assumptions heuristically extracted on user interaction. The presented strategies have been also combined with features filtering within object representation. The filtering includes grouping contextualizing and varying features set representations.

6.4. IKONA/MAESTRO sofware

Participants: Vera Bakic, Laurent Joyeux, Sofiene Mouine, Souheil Selmi.

This year, IKONA has been extended in the context of Pl@ntNet, Glocal and I-SEARCH projects.

- For the Pl@ntNet project, along the continuing improvements and optimizations in the MAESTRO software, a number of new features were added:

new options for interest points distribution and filtering with the segmented image,
a new shape context descriptor (corresponding to [11]),
various combinations of descriptors in one vector per interest point or region,
for regions: extraction of sub-images, EOH and Fourier descriptors,
more options for update of calculated signatures, new score type (used in ImageCLEF2012) and decision rules (adaptive Knn calculation based on individual plant information) for statistical tools.
In addition, a number of new web services and functionality were developed/updated and deployed:
the addition of new databases (Vignes, Musa), while some with the organ annotation (Photoflora, Girod), and the update of the multi search views for the new datasets;
the development of new services allowing to return of botanical information in several formats (csv, xml, JSON...);
the update of the indexation system so that it can crawl images from different sources (internal or external sources like CEL web service, which uses html or identiplante web service (in JSON format);
the development of html pages to annotate a set of images by organ type;
the development of an applet demonstration of leaf architecture (this applet interfaces a library developed by a PhD student working on this project);
the development of an application "Pl@ntNet Identify" for android platform and its interfacing with the existing web services.

- For the Glocal project, functionalities such as fraud detection and similarity search were integrated in the mock up of user interface and in the final demonstration of the project. In addition, a dynamic indexation system of images from AFP (Agence France-Presse) was implemented as well as the similarity web services working on this dynamic dataset.
- For the I-SEARCH project, an integration of the video mining component in I-SEARCH platform was done. The component extracts visual objects that are the most recurrent from a set of images, or in a video.

IN-SITU Project-Team

6. New Results

6.1. Interaction Techniques

Participants: Caroline Appert, Michel Beaudouin-Lafon, David Bonnet, Anastasia Bezerianos, Olivier Chapuis, Emilien Ghomi, Stéphane Huot, Can Liu, Wendy Mackay [correspondant], Mathieu Nancel, Cyprien Pindat, Emmanuel Pietriga, Theophanis Tsandilas, Julie Wagner.

We explore interaction techniques in a variety of contexts, including individual interaction techniques on mobile devices, the desktop, and very large wall-sized displays, using one or both hands. We also explore interaction with physical objects and across multiple devices, to create mixed or augmented reality systems. This year, we explored interaction techniques based on time (*EWE* and *Dwell-and-Spring*), bimanual interaction on mobile devices (*Bipad*) and interaction on very large wall displays (*Jelly Lenses*, *Looking-Around-Bezels*). We also developed interactive paper systems to support early, creative design (*Pen-based Mobile Assistants*, *Paper Tonnetz*, *Paper Substrates*). We also explored augmented reality systems, using tactile feedback (*TactileSnowboard Instructions*) and tangible interaction (*Mobile AR*, *Combinatorix*) to support learning.

EWE – Although basic interaction techniques, such as multiple clicks or spring-loaded widgets, take advantage of the temporal dimension, more advanced uses of rhythmic patterns have received little attention in HCI. Using temporal structures to convey information can be particularly useful in situations where the visual channel is overloaded or even not available. We introduce Rhythmic Interaction [24] which uses rhythm as an input technique (Figure 10). Two experiments demonstrate that (i) rhythmic patterns can be efficiently reproduced by novice users and recognized by computer algorithms, and (ii) rhythmic patterns can be memorized as efficiently as traditional shortcuts when associated with visual commands. Overall, these results demonstrate the potential of Rhythmic Interaction and richer repertoire of interaction techniques. (*Best Paper award, CHI'12*)

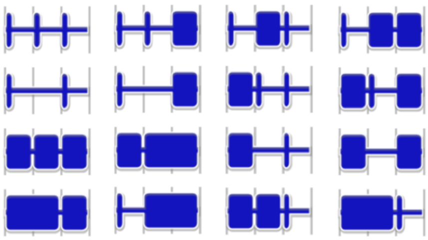


Figure 10. We defined 16 three-beat patterns: Each rectangle represents a tap, the thin gray lines show beats.

Dwell-and-Spring – Direct manipulation interfaces consist of incremental actions that should be reversible The challenge is how to provide users with an effective "undo". Actions such as manipulating geometrical shapes in a vector graphics editor, navigating a document using a scrollbar, or moving and resizing windows on the desktop rarely offer an undo mechanism. Users must manually revert to the previous state by recreating a similar sequence of direct manipulation actions, with a high level of associated motor and cognitive costs. We need a consistent mechanism that supports undo in multiple contexts. *Dwell-and-Spring* [19] uses a spring metaphor that lets users undo a variety of direct manipulations of the interface. A spring widget pops up whenever the user dwells during a press-drag-release interaction, giving her the opportunity to either cancel the current manipulation (Figure 11) or undo the last one. The technique is generic and can easily be implemented on top of existing applications to complement the traditional undo command. A controlled experiment demonstrated that users can easily discover the technique and adopt it quickly when it is discovered.

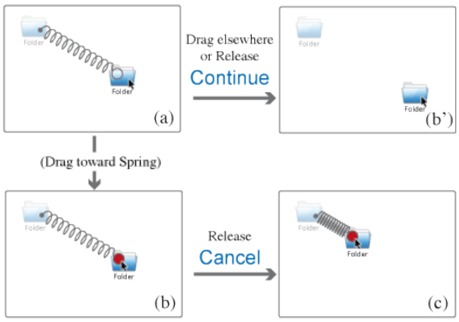


Figure 11. Cancel scenario: The user dwells while dragging an icon (a), which pops up a spring. She either (b) catches the spring's handle and releases the mouse button to cancel the current drag and drop, causing the spring to shrink smoothly (c) and returning the cursor and icon to their original locations, or she continues dragging the spring's handle any direction (b').

BiPad – Although bimanual interaction is common on large tabletops, it is rare on hand-held devices. We take advantage of the advanced multitouch input capabilities available on today's tablets to introduce new bimanual interactio techniques, under a variety of mobility conditions. We found that, when users hold a tablet, the primary function of the non-domininant hand is to provide support, which limits its potential movement. We studied how users "naturally" hold multi-touch tablets to identify comfortable holds, and then developed a set of 10 two-handed interaction techniques that accounts for the need to support the device while interacting with it. We introduced the *BiTouch* design space that extends Guiard's "Kinematic Chain Theory" [49] to account for the *support function* in bimanual interaction. We also designed and implemented the *BiPad* toolkit and set of interactions, which enables us to implement bimanual interaction on multitouch tablets (Figure 12). Finally, a controlled experiment demonstrated the benefits and trade-offs among specific techniques and offered insights for designing bimanual interaction on hand-held devices [31].

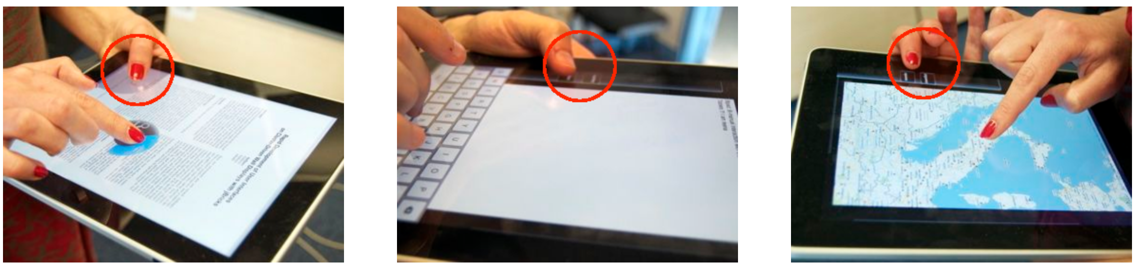


Figure 12. Bimanual interaction on a multitouch tablet with BiPad: (left) navigating in a document; (center) switching to uppercase while typing on a virtual keyboard; (right) zooming a map. The non-dominant hand is holding the device and could perform 'tapa', 'gestures' or 'chords' in order to augment dominant hand's interactions.

Jelly Lenses – Focus+context lens-based techniques smoothly integrate two levels of detail using spatial distortion to connect the magnified region and the context. Distortion guarantees visual continuity, but causes problems of interpretation and focus targeting, partly due to the fact that most techniques are based on statically-defined, regular lens shapes, that result in far-from-optimal magnification and distortion (Figure 13 left and center). JellyLenses [27] dynamically adapt to the shape of the objects of interest, providing detail-in-context visualizations of higher relevance by optimizing what regions fall into the focus, context and spatially-distorted transition regions (Figure 13 -right). A multi-scale visual search task experiment demonstrated that JellyLenses consistently perform better than regular fisheye lenses.

Figure 13. Magnifying the Lido in Venice. (left) a small fisheye magnifies one part of the island (Adriatic sea to Laguna Veneta), but requires extensive navigation to the whole island in detail; (center) a large fisheye magnifies a bigger part of the island, but severely distorts almost the entire image, hiding other islands; (right) a JellyLens automatically adapts its shape to the region of interest, with as much relevant information in the focus as (b) while better preserving context: surrounding islands are almost untouched from (a).

Looking behind Bezels – Using tiled monitors to build wall-sized displays has multiple advantages: higher pixel density, simpler setup and easier calibration. However, the resulting display walls suffer from the visual discontinuity caused by the bezels that frame each monitor. To avoid introducing distortion, the image has to be rendered as if some pixels were drawn behind the bezels. In turn, this raises the issue that a non-negligible part of the rendered image, that might contain important information, is visually occluded. We drew upon the analogy to french windows that is often used to describe this approach, and make the display really behave as if the visualization were observed through a french window [21]. We designed and evaluated two interaction techniques that let users reveal content hidden behind bezels. One enables users to offset the entire image through explicit touch gestures. The other adopts a more implicit approach: it makes the grid formed by bezels act like a true french window using head tracking to simulate motion parallax, adapting to users' physical movements in front of the display. The two techniques work for both single- and multiple-user contexts.

Pen-based Mobile Assistants – Digital pen technology allows easy transfer of pen data from paper to the computer. However, linking handwritten content with the digital world remains difficult as it requires the translation of unstructured and highly personal vocabularies into data structured so as to be understood and processed by a computer. Automatic recognition can help, but is not always reliable: it require active cooperation between users and recognition algorithms. We examined [30] the use of portable touch-screen devices in connection with pen and paper to help users direct and refine the interpretation of their strokes on paper. We explored four bimanual interaction techniques that combine touch and pen-writing, where user attention is divided between the original strokes on paper and their interpretation by the electronic device. We demonstrated these techniques through a mobile interface for writing music (Figure 14) that complements the automatic recognition with interactive user-driven interpretation. An experiment evaluated the four techniques and provided insights as to their strengths and limitations.

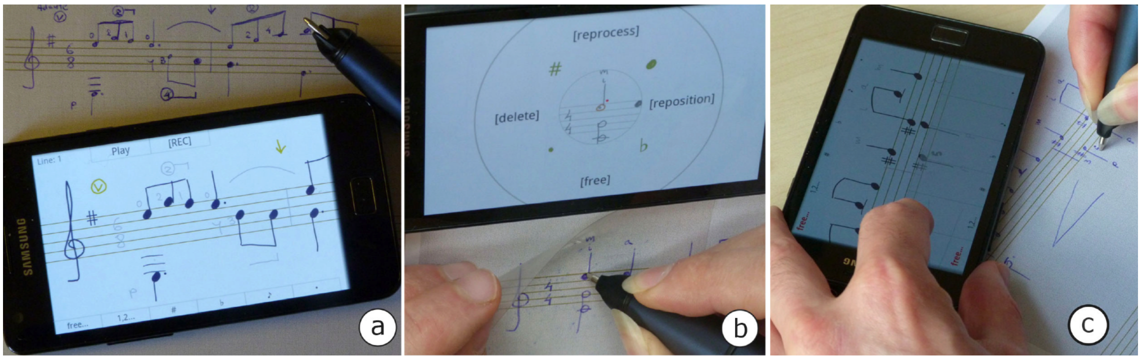


Figure 14. Writing music with a pen and a smartphone. (a) Handwritten score translated by the device. (b) Correcting the recognition of a note over a plastic sheet. (c) Guiding the interpretation of strokes with the left hand.

Paper Tonnetz – Tonnetz are space-based musical representations that lay out individual pitches in a regular structure. We investigated how properties of Tonnetz can be applied in the composition process, including how to represent pitch based on chords or scales and lay them out in a two-dimensional space (Figure 15). *PaperTonnetz*[20] is a tool that lets musicians explore and compose music with Tonnetz representations by making gestures on interactive paper, creating replayable patterns that represent pitch sequences and/or chords. An initial test in a public setting demonstrated significant differences between novice and experienced musicians and led to a revised version that explicitly supports discovering, improvising and assembling musical sequences in a Tonnetz.

Paper Substrates – Our goal is to design novel interactive paper interfaces that support the creative process. We ran a series of participatory design sessions with music composers to explore the concept of “paper substrates” [23]. Substrates are ordinary pieces of paper, printed with an Anoto dot pattern, that suppport a variety of advanced forms of interaction (Figure 15). Each substrate is strongly typed, such as a musical score or a graph, which faciliates interpretation by the computer. The composers were able to create, manipulate and combine layers of data, rearranging them in time and space as an integral part of the creative process. Moreover, the substrates approach fully supported an iterative process in which templates can evolve and be reused, resulting in highly personal and powerful interfaces. We found that paper substrates take on different roles, serving as data containers, data filters, and selectors. The design sessions resulted in several pen interactions and tangible manipulations of paper components to support these roles: drawing and modifying specialized data over formatted paper, exploring variations by superimposing handwritten data, defining programmable modules, aligning movable substrates, linking them together, overlaying them, and archiving them into physical folders.

Tactile Snowboard Instructions – Beginnning snowboarders have difficulty getting instructions and feedback on their performance if they are separated spatially from their coach. Snowboarders can learn correct technique by wearing a system with actuators (vibration motors) attached to the thighs and shoulders, which reminds them to shift their weight and to turn their upper body in the correct direction (Figure 16). A field study with amateur snowboarders demonstrated that these “tactile instructions” are effective for learning basic turns and offered recommendations for incorporating tactile instructions into sports training. *Best Paper award, Mobile HCI’12*

Mobile AR – We examined how new capabilities of hand-held devices, specifically higher resolution screens, camera and localization, can be used to create mobile *Augmented Reality (AR)* to help users learn and manage their interactions with everyday physical objects, such as door codes and home appliances. We explored AR-based mobile note-taking [50] to provide real-time on-screen feedback of physical objects that the user must manipulate, such as entering a door code. Here, the user uses the device to identify the required values of sliders

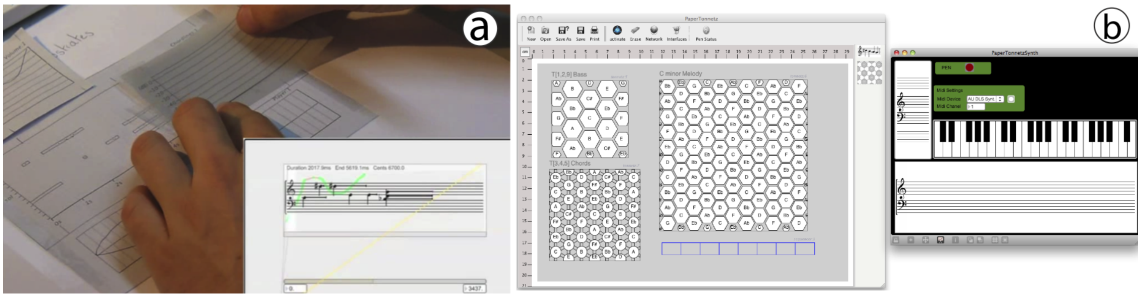


Figure 15. Paper-based interfaces for musical creation. (a) Paper substrates are interactive paper components for working with musical data. (b) PaperTonnetz main interface representing the virtual page with three Tonnetz and one sequencer (left). The Max/MSP patch to play and visualize created sequences (right).

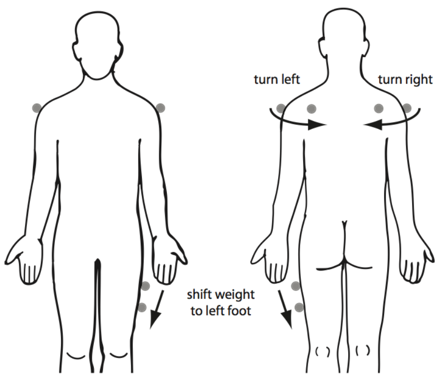


Figure 16. Two vibration motors are placed at each shoulder and laterally at the thigh that points forward during the ride. Arrows illustrate the direction of the stimuli on the skin, labels show the corresponding messages.

and buttons (Figure 17). A controlled experiment showed that mobile AR improved both speed and accuracy over traditional text or picture-based instructions. We also demonstrated that adding real-time feedback in the AR layer that shows the user's actions with respect to the physical controls further increases performance [25]. (*Honorable Mention award, CHI 2012*)

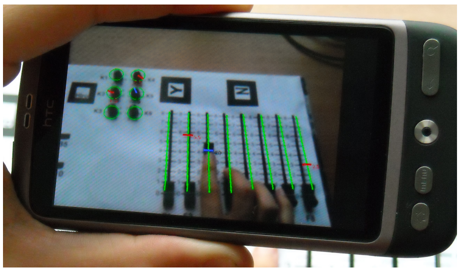


Figure 17. Mobile augmented reality for setting physical controls. Required values are displayed in red and turn blue when set correctly.

Combinatorix – We developed Combinatorix [28], a mixed tabletop system to help groups of students work collaboratively to solve probability problems. Users combine tangible objects in different orders and watch the effects of various constraints on the problem space (Figure 18). A second screen displays an abstract representation, such as a probability tree, to show how their actions influenced the total number of combinations. We followed an iterative participatory design process with college students taking a combinatorics class and demonstrated the benefits of using a tangible approach to facilitate learning abstract concepts.

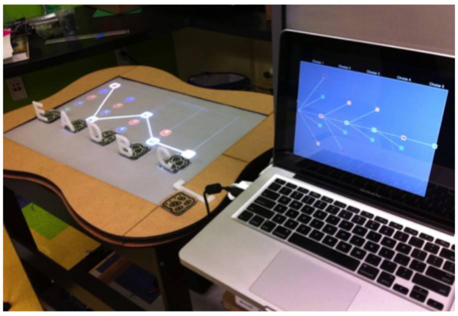


Figure 18. Combinatorix uses tangible objects on an interactive tabletop to control the tabletop display and associated screen, to help users explore and understand complex problems in combinatorial statistics.

6.2. Research Methods

Participants: Caroline Appert, Michel Beaudouin-Lafon, Anastasia Bezerianos, Olivier Chapuis, Jérémie Garcia, Stéphane Huot, Ilaria Liccardi, Wendy Mackay [correspondant], Emmanuel Pietriga.

Human-Computer Interaction is a multi-disciplinary field, with elements of computer science, software engineering, experimental psychology and anthropology. More recently, designers have joined the CHI community, offering an important perspective, but also a different fundamental research paradigm, which differs from the value systems of engineering and the natural sciences. We explored the paradigm of *Research through Design* [34], which we differentiate from traditional epistemologies in the human sciences. We distinguish design from research-through-design: the end goal is not to produce an artifact, but rather to frame an alternative future and uncover unmet human needs, desires, emotions, and aspirations. We identified three research perspectives that have been adopted within the HCI community: *Projection* explores possible future states, *Place* specifies the context in which design artifacts presented to gather data, and *Point-of-View* identifies the philosophical perspective imposed by researchers. Our goal is to understand what it means to conduct research through design and how to value research-through-design contributions.

In addition to exploring general questions about research paradigms, we also explore more focused questions that apply new research methods to the design of multi-surface environments. Large interactive surfaces (like the WILD platform) are interesting collaborative settings that allow viewing of large amounts of visual information. Before we are ready to use these platforms in real visual analysis situations, where people can place themselves at different positions around the display, we must first understand how the perception of visual information is affected by perspective distortion introduced by varying viewing distances and angles. A deeper understanding of such distortion effects can help visualization researchers design effective visualizations for these spaces and implement interaction techniques to aid in extreme distortion situations. We conducted [16] two studies on the perception of visual variables that encode information, such as Angle, Area, and Length, and found that perception is impacted differently at different locations on the screen, depending on the vertical and horizontal positioning of this information. The visual variable of Length was the most accurately perceived across the display. Our second study examined the effect of perception when participants can move freely in such situations, compared to when they have a static viewpoint, and found that a far but static viewpoint was as accurate but less time consuming than one that included free motion. But we observed that in free motion participants often chose non-optimal walking strategies that can increased perception errors, thus we provide precise recommendations on where and how to move in such environments. This work is a first step towards understanding and predicting the impact of large display environments to people's understanding and tasks.

Annotations play an important role in visual analysis and record-keeping. We discuss the use of annotations on visualization dashboards citeelias:hal-00719221, collections of linked visualizations, focusing on business intelligence analysis through a user centered design process from expert analysts in this domain. The first contribution bridges the gap between expert analyst needs and designers, when it comes to visualization annotations. The second offers a novel approach to annotating of visualizations, "context aware annotations": We annotate data queries directly, rather than image/chart locations. Annotations are present irrespective of the visual data representations users select (different charts, numeric tabular views of their data, etc). We focus particularly on novel annotation aspects made possible by this approach, such as multi-visualization annotations, annotation transparency across charts and data hierarchy levels, and the possibility of recommending to users annotations done to similar data to enable annotation re-use. We also consider new challenges that arise from such our approach, such as what happens to annotations when the underlying data is changed, and provide recommendations and design solutions.

6.3. Engineering of interactive systems

Participants: Caroline Appert, Michel Beaudouin-Lafon [correspondant], Olivier Chapuis, Stéphane Huot, Wendy Mackay, Emmanuel Pietriga, Clément Pillias, Romain Primet.

INSITU continues to develop and apply toolkits to explore and implement interactive systems. Most of the projects listed in the *Interaction Techniques* section either build upon existing toolkits, e.g., *Jelly Lenses* to improve management of focus+context on a wall-sized display, or created new ones, e.g., *BiPad* to create various bimanual interaction techniques for hand-held tablets.

INSITU's primary testbed for exploring multi-surface interaction is the WILD Room [15] (Wall-sized Interaction with Large Datasets), a multisurface environment featuring a wall-sized display, a multitouch table, and various mobile devices. Our goal is to explore the next generation of interactive systems by distributing interaction across these diverse computing devices, enabling multiple users to easily and seamlessly create, share, and manipulate digital content. Our research strategy is to design an extreme environment that pushes the limits of technology – both hardware and software. To ground the design process, we work with extreme users – scientists whose daily work both inspires and stress-tests the environment as they seek to understand exceptionally large and complex datasets. The WILD room, and the soon-to-be-built WILDER room are part of *DigiScope*, a 22.5 Meuro “Equipement d’Excellence” project led by INSITU.

INSITU’s collaboration with the ALMA radio-telescope on the design and implementation of user interfaces for operations monitoring and control continued this year [26], and was eventually transferred to Inria Chile in July (see Section 7.4.2). The ALMA radio-telescope, currently under construction in northern Chile, is a very advanced instrument that presents numerous challenges. From a software perspective, one critical issue is the design of graphical user interfaces for operations monitoring and control that scale to the complexity of the system and to the massive amounts of data users are faced with. Early experience operating the telescope with only a few antennas showed that conventional, WIMP-based user interfaces are not adequate in this context. They consume too much screen real-estate, require many unnecessary interactions to access relevant information, and fail to provide operators and astronomers with a clear mental map of the instrument. They increase extraneous cognitive load, impeding tasks that call for quick diagnosis and action. To address this challenge, the ALMA software division adopted a user-centered design approach in collaboration with members of INSIUT. For the last two years, astronomers, operators, software engineers and human-computer interaction researchers from INSITU have been working on the design of better user interfaces based on state-of-the-art visualization techniques. This eventually led to the joint development of those interface components using various software toolkits, some of them developed at INSITU (Section 5.2).

LAGADIC Project-Team

6. New Results

6.1. Visual tracking

6.1.1. 3D model-based tracking

Participants: Antoine Petit, Eric Marchand.

Our 3D model-based tracking algorithm [2] was used in various contexts. We began a collaboration with Astrium EADS in 2010 in order to build a more versatile algorithm able to consider complex objects. The main principle is to align the projection of the 3D model of the object with observations made in the image for providing the relative pose between the camera and the object using a non-linear iterative optimization method. The approach proposed takes advantage of GPU acceleration and 3D rendering. From the rendered model, visible edges are extracted, from both depth and texture discontinuities. Potential applications would be the final phase of space rendezvous mission, in-orbit servicing, large debris removal using visual navigation, or airborne refuelling [41], [40], [32].

6.1.2. Omnidirectional vision system

Participant: Eric Marchand.

In this study performed in collaboration with Guillaume Caron and El Mustapha Mouaddib from Mis in Amiens, we have been interested by the redundancy brought by stereovision in omnidirectional vision sensors. This has been obtained by combining a single camera and multiple mirrors. Within this framework, we proposed to extend the 3D model-based tracking algorithm [2] for such system [15].

Thanks to a collaboration with Esiea in Laval, France, and the Inria and Irisa Hybrid team, we developed a system named Flyviz that has been patented. It is composed of a helmet mounted catadioptric camera coupled with an immersive display. The image acquired by the sensor is processed to give the user a full 360-degree panoramic view [27].

6.1.3. Pose estimation using mutual information

Participant: Eric Marchand.

Our work with Amaury Dame related to template tracking using mutual information [17] as registration criterion has been extended to 3D pose estimation using a 3D model. Since a homography was estimated, the tracking approach presented in [17] was usable for planar scenes. The new approach [45] can be considered for any scene or camera motion. Considering mutual information as similarity criterion, this approach is robust to noise, lighting variations and does not require a statistically robust estimation process. It has been used for visual odometry in large scale environment.

6.1.4. Pseudo-semantic segmentation

Participants: Rafik Sekkal, François Pasteau, Marie Babel.

To address the challenge of tracking initialization issues, we investigate joint segmentation and tracking approaches characterized by resolution and hierarchy scalability as well as a low computational complexity. Through an original scalable Region Adjacency Graph (RAG), regions can be adaptively processed at different scale representations according to the target application [42]. The results of this pseudo-semantic segmentation process are further used to initialize the object tracker (patch, visual objects, planes...) on several scales of resolutions.

6.1.5. Augmented reality using RGB-D camera

Participants: Hideaki Uchiyama, Eric Marchand.

We consider detection and pose estimation methods of texture-less planar objects using RGB-D cameras. It consists in transforming features extracted from the color image to a canonical view using depth data in order to obtain a representation invariant to rotation, scale, and perspective deformations. The approach does not require to generate warped versions of the templates, which is commonly needed by existing object detection techniques [35].

We also investigate the use of RGB-D sensors for object detection and pose estimation from natural features. The proposed method exploits depth information to improve keypoint matching of perspectively distorted images. This is achieved by generating a projective rectification of a patch around the keypoint, which is normalized with respect to perspective distortions and scale [34].

6.2. Visual servoing

6.2.1. Visual servoing using the sum of conditional variance

Participants: Bertrand Delabarre, Eric Marchand.

Within our study of direct visual servoing, we propose a new similarity function: the use of the sum of conditional variance [31] that replace SSD or mutual information [3]. It has been shown to be invariant to non-linear illumination variations and inexpensive to compute. Compared to other direct approaches of visual servoing, it is a good trade off between techniques using the pixels luminanc,e which are computationally inexpensive but non robust to illumination variations, and other approaches using the mutual information, which are more complicated to compute but offer more robustness towards the variations of the scene.

6.2.2. Photometric moment-based visual servoing

Participants: Manikandan Bakthavatchalam, Eric Marchand, François Chaumette.

The direct visual servoing approaches that have been developed in the group in the recent years, either using the luminance of each pixel, or the mutual information [3], or the sum of conditional variance described just above, allows reaching an excellent positioning accuracy. This good property is however counterbalanced by a small convergence domain due to the strong non linearities involved in the control scheme. To remedy to these problems, we started a study on using photometric moments as inputs of visual servoing. We expect to find again the nice decoupling and large convergence domain that we obtained for binary moments, without the need of any object segmentation.

6.2.3. Visual servoing using RGB-D sensors

Participants: Céline Teulière, Eric Marchand.

We propose a novel 3D servoing approach [43] that uses dense depth maps to perform robotic tasks. With respect to pose-based approaches, our method does not require the estimation of the 3D pose, nor the extraction and matching of 3D features. It only requires dense depth maps provided by 3D sensors. Our approach has been validated in servoing experiments using the depth information from a low cost RGB-D sensor. Thanks to the introduction of M-estimator in the control law, positioning tasks are properly achieved despite the noisy measurements, even when partial occlusions or scene modifications occur.

6.2.4. Visual servoing of cable-driven parallel robot

Participant: François Chaumette.

This study is realized in collaboration with Rémy Ramadour and Jean-Pierre Merlet from EPI Coprin at Inria Sophia Antipolis. Its goal is to adapt visual servoing techniques for cable-driven parallel robot in order to achieve acurate manipulation tasks. This study is in the scope of the Inria large-scale initiative action Pal (see Section 8.2.7).

6.2.5. Micro-Nanomanipulation

Participants: Eric Marchand, Le Cui.

In collaboration with Femto-ST in Besançon, we developed an accurate nanopositioning system based on direct visual servoing [20]. This technique relies only on the pure image signal to design the control law, by using the pixel intensity of each pixel as visual features. The proposed approach has been tested in terms of accuracy and robustness in several experimental conditions. The obtained results have demonstrated a good behavior of the control law and very good positioning accuracy: 89 nm, 14 nm, and 0.001 degrees in the x, y and θ_z axes of a positioning platform, respectively.

We begin a work, within the ANR P2N Nanorobust project (see Section 8.2.4), on the development of micro- and nano-manipulation within SEM (Scanning Electron Microscope). Our goal is to provide visual servoing techniques for positioning and manipulation tasks with a nanometer precision.

6.2.6. *Autonomous landing by visual servoing*

Participants: Laurent Coutard, François Chaumette.

This study was realized in collaboration with Dassault Aviation with the financial support of DGA. It was concerned with the autonomous landing of fixed wing aircrafts on carrier by visual servoing. A complete system has been developed [12]. The vision part consists in detecting the carrier in the image sequence and then tracking it using either dense template tracking or our 3D model-based tracker [2]. The visual servoing part consists in computing particular visual features able to correctly handle the aircraft degrees of freedom. Perturbations due to the wind and carrier motions have also been considered. The complete system has been validated in simulation using synthetic images provided by Xplane simulator and a dynamic model of the aircraft provided by Dassault Aviation.

6.3. Visual navigation of mobile robots

6.3.1. *Visual navigation using mutual information*

Participants: Eric Marchand, Bertrand Delabarre.

We have developed a visual navigation scheme based on the mutual information between the images acquired by an onboard camera and a visual memory to control the orientation of a vehicle during its navigation [18].

We also proposed to extend this approach to visual servoing with vision systems that consider the unified sphere model for central cameras using a normalized version of the mutual information. This permitted to apply the technique to large fields of view with a more reliable similarity function [30].

6.3.2. *3D Mapping and real time navigation*

Participants: Maxime Meilland, Patrick Rives.

This study was realized in collaboration with Andrew Comport from I3S in Sophia Antipolis. Our approach relies on a monocular camera on board the vehicle and the use of a database of spherical images of the scene acquired during an offline step [14]. This geo-referenced database allows us to obtain a robust **drift free** localization. Basically, the database is composed of spherical images augmented by depth that are positioned in a GIS (Geographic information system). This spherical robot centered representation accurately represents all necessary information for vision-based navigation and mapping [37]. During the online navigation, the vehicle pose is computed by aligning the current image acquired by the camera with the closest reference sphere extracted from the database [26].

6.3.3. *Indoors Slam*

Participants: Cyril Joly, Patrick Rives, Pierre Martin, Eric Marchand.

We developed in Sophia Antipolis a new Slam method fusing laser scan data with the spherical images provided by an omnidirectional camera. Thanks to the trace of the laser scan projected onto the spherical view, we are able to compute a RGB-D model of the environment by using a dense visual Slam approach.

In Rennes and in collaboration with Orange Labs, we considered the development of a visual Slam algorithm. Since the targeted platforms in this this study are Android Smartphone, sequential Slam approaches have been studied.

6.3.4. Topological navigation

Participants: Alexandre Chapoulie, Patrick Rives.

This study is realized in collaboration with David Filliat from Ensta in Paris. Navigation algorithms are often sensitive to the robot orientation involving an impossibility to detect a place already visited from a different point of view. In order to alleviate this drawback, panoramic or omnidirectional cameras are often used. We have developed a loop closure detection algorithm based on an ego-centric spherical view that satisfies, in addition to other properties, a robot orientation independence [11].

A topological model captures the accessibility of the different places in the environment and allows a coarse localization. From a sequence of spherical views, we have developed a context-based segmentation algorithm. We hence define a topological place as having a structure which does not change, variation leading to a place change. The structure variations are detected with an efficient change-point detection algorithm [28].

6.3.5. Development of an autonomous shopping cart

Participants: Luca Marchetti, Patrick Rives.

This work is realized in collaboration with Pascal Morin from Isir in Paris. It consists in developing a shopping cart with autonomy capabilities (automatic user following, obstacle avoidance, etc), as part of the Inria Large-scale initiative action Pal, which aims at developing robotic tools for disabled persons or elderlies (see Section 8.2.7). Experiments have been successfully conducted both on the mobile robot Hannibal and on the wheeled walking aid ANG (Assistive Navigation Guide) developed by the EPI Coprin in Sophia Antipolis [36].

6.3.6. Automous navigation of wheelchairs

Participants: Rafik Sekkal, François Pasteau, Marie Babel.

This study is aimed at designing a robotic vision-based system dedicated to assisted navigation of electrical wheelchair in an unkown environment. In particular, going through doors, taking the elevator in a secure way without risking collision because of hazardous wheelchair motions remain a relevant issue. The idea is here to provide an embedded and flexible system able to ensure the immediate compatibility of the proposed system with existing electrical wheelchairs. From the platform described in Section 5.5 , we first addressed the door detection issue for automatically initializing the tracking process that is required for localisation and navigation purposes. We then defined a low complex solution of automatic door recognition that can be decomposed into three successive steps: line extraction (LSD-based algorithm), vanishing point estimation and door recognition itself by using geometrical cues. As soon as a door is detected and tracked through model-based trackers, the idea is to take into account the position of the wheelchair joystick in order to interpret the intention of the user. First experiments have shown the validity of the proposed approach. This study is conducted in conjunction with the scope of the Inria large-scale initiative action Pal (see Section 8.2.7).

6.3.7. Obstacle avoidance

Participants: Fabien Spindler, François Chaumette.

This study was realized in collaboration with Andrea Cherubini who is now Assistant Prof. at Université de Montpellier. It is concerned with our long term researches about visual navigation from a visual memory without any accurate 3D localization [9]. In order to deal with obstacle avoidance while preserving the visibility in the visual memory, we have proposed a control scheme based on tentacles for fusing the data provided by a pan-tilt camera and a laser range sensor [16].

6.4. Medical robotics

6.4.1. Visual servoing based on dense ultrasound information

Participants: Caroline Nadeau, Alexandre Krupa.

In the context of the ANR USComp project (see Section 8.2.3), we pursued our works on the development of ultrasound image-based visual servoing methods that directly use pixel intensities of the ultrasound image as control inputs. In opposite with methods based on geometrical visual features, this new approach does not require any image segmentation step that is difficult to robustly perform on ultrasound images. By coupling our method with a predictive control law based on the periodicity of physiological motion, we propose a solution to stabilize the ultrasound image by actively compensating the physiological motions of the patient. The principle consists in automatically synchronizing the 6 DOF motion of a 2D or 3D probe with the rigid motion of a soft tissue target. First ex-vivo results obtained on animal tissues demonstrated the validity of the concept [39].

In collaboration with Prof. Pierre Dupont from Harvard University at Boston, we also addressed the motion tracking of a target that can consist of either the tip of a robot inserted on a beating heart or cardiac tissues. Unlike the previous work, where the motion compensation task was realized physically by moving the probe attached to a robotic arm, we propose here to track the motion of the target using a 3D region of interest (ROI) which is automatically moved within the whole volume observed by a 3D probe thanks to our intensity-based ultrasound visual servoing method. In vivo animal experiments were conducted in Children's Hospital at Boston and validated this tracking approach [38].

6.4.2. Autonomous control modes for ultrasound probe guidance

Participants: Tao Li, Alexandre Krupa.

In the context of the ANR Prosit project (see Section 8.2.2), we proposed several autonomous control modes in order to assist a doctor during a robotized and teleoperated ultrasound examination (tele-echography). This year we developed an assistance functionality that automatically maintains the visibility of an anatomic element of interest while the doctor teleoperates the 2D ultrasound probe held by the medical robot. The method is based on a multi-task controller that gradually activates an ultrasound visual servoing in case some geometrical features leave a pre-defined safe area of the image in order to bring them back inside the view [33]. With this approach the DOFs of the robotized probe are not exclusively constrained by the visibility task but also available for the tele-operation. This new assistance functionality was implemented on the ANR Prosit robotic platform and first in vivo results obtained on a human volunteer validated the concept.

6.4.3. Real-time soft-tissue deformation tracking in 3D ultrasound

Participant: Alexandre Krupa.

We proposed a dense ultrasound tracking algorithm that estimates in real time both rigid and non-rigid motions of a region of interest observed in a sequence of 3D ultrasound images. The deformation is modeled by 3D thin-plate splines (TPS) whose parameters are estimated online from intensity difference measured in successive volumes. To increase the robustness of this approach to image noise, we proposed two solutions to mechanically constrain the deformable model. The first is based on the addition of a regularization term in the TPS model and the second consists in coupling the TPS with a mass-spring system. These methods were validated on simulated sequences of deformed 3D ultrasound images.

6.4.4. Needle detection and tracking in 3D ultrasound

Participant: Alexandre Krupa.

We designed an algorithm able to detect a needle inserted manually in a 3D ultrasound volume from an arbitrary point, and able to robustly track this needle in real-time. We also experimentally demonstrated the possibility to guide the ultrasound probe to keep the needle visible and aligned, using visual servoing. Such a system could assist an operator during manual insertions, which are currently performed under free-hand ultrasound monitoring. In addition, we plan in future works to combine this method to a needle steering robotic system for guiding accurately the needle toward a target while optimizing its visibility.

LEAR Project-Team

6. New Results

6.1. Visual recognition in images

6.1.1. Correlation-Based Burstiness for Logo Retrieval

Participants: Matthijs Douze, Jerome Revaud, Cordelia Schmid.

Detecting logos in photos is challenging. A reason is that logos locally resemble patterns frequently seen in random images. In [21] we propose to learn a statistical model for the distribution of incorrect detections output by an image matching algorithm. It results in a novel scoring criterion in which the weight of correlated keypoint matches is reduced, penalizing irrelevant logo detections. In experiments on two very different logo retrieval benchmarks, our approach largely improves over the standard matching criterion as well as other state-of-the-art approaches.

Figure 1. Illustration of a logo detected by our method.

6.1.2. Towards Good Practice in Large-Scale Learning for Image Classification

Participants: Zeynep Akata, Zaid Harchaoui, Florent Perronnin [XRCE], Cordelia Schmid.

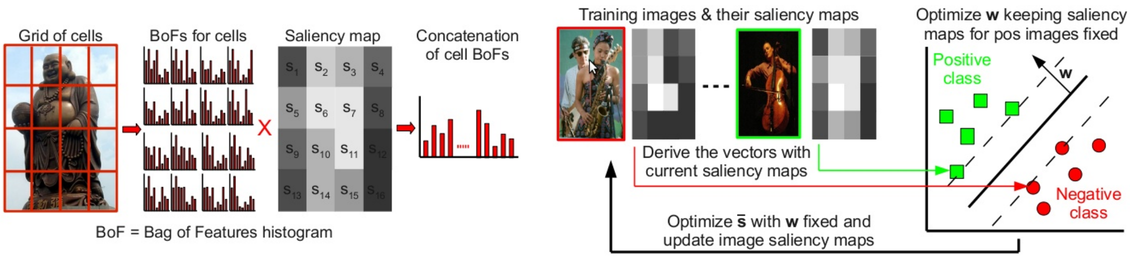
In [19] we propose a benchmark of several objective functions for large-scale image classification: we compare the one-vs-rest, multiclass, ranking and weighted average ranking SVMs. Using stochastic gradient descent optimization, we can scale the learning to millions of images and thousands of classes. Our experimental evaluation shows that ranking based algorithms do not outperform a one-vs-rest strategy and that the gap between the different algorithms reduces in case of high-dimensional data. We also show that for one-vs-rest, learning through cross-validation the optimal degree of imbalance between the positive and the negative samples can have a significant impact. Furthermore, early stopping can be used as an effective regularization strategy when training with stochastic gradient algorithms. Following these “good practices”, we were able to improve the state-of-the-art on a large subset of 10K classes and 9M of images of ImageNet from 16.7% accuracy to 19.1%. Some qualitative results can be seen in Figure 2 .

6.1.3. Discriminative Spatial Saliency for Image Classification

Participants: Frédéric Jurie [Université de Caen], Cordelia Schmid, Gaurav Sharma.

Figure 2. ImageNet10K results (top-1 accuracy in %) obtained with w-OVR and 130K-dim Fisher vectors. (a-d) Sample classes among the best performing ones. (e-h) Sample classes among the worst performing ones.

In many visual classification tasks the spatial distribution of discriminative information is (i) non uniform e.g. "person reading" can be distinguished from "taking a photo" based on the area around the arms i.e. ignoring the legs, and (ii) has intra class variations e.g. different readers may hold the books differently. Motivated by these observations, we propose in [22] to learn the discriminative spatial saliency of images while simultaneously learning a max-margin classifier for a given visual classification task. Using the saliency maps to weight the corresponding visual features improves the discriminative power of the image representation. We treat the saliency maps as latent variables and allow them to adapt to the image content to maximize the classification score, while regularizing the change in the saliency maps. See Figure 3 for an illustration. Our experimental results on three challenging datasets, for (i) human action classification, (ii) fine grained classification, and (iii) scene classification, demonstrate the effectiveness and wide applicability of the method.



(a) (b)

Figure 3. (a) The images are represented by concatenation of cell bag-of-features weighted by the image saliency maps. (b) We propose to use a block coordinate descent algorithm for learning our model. As in a latent SVM, we optimize in one step the weight vector w keeping the saliency maps of the positive images fixed, and in the other step we optimize the saliency keeping w fixed.

6.1.4. Tree-structured CRF Models for Interactive Image Labeling

Participants: Gabriela Csurka [XRCE], Thomas Mensink, Jakob Verbeek.

In [8] we propose structured prediction models for image labeling that explicitly take into account dependencies among image labels. In our tree structured models, image labels are nodes, and edges encode dependency relations. To allow for more complex dependencies, we combine labels in a single node, and use mixtures of trees. Our models are more expressive than independent predictors, and lead to more accurate label predictions. The gain becomes more significant in an interactive scenario where a user provides the value of some of the image labels at test time. Such an interactive scenario offers an interesting trade-off between label accuracy and manual labeling effort. The structured models are used to decide which labels should be set by the user, and transfer the user input to more accurate predictions on other image labels. We also apply our models to attribute-based image classification, where attribute predictions of a test image are mapped to class probabilities by means of a given attribute-class mapping. Experimental results on three publicly available benchmark data sets show that in all scenarios our structured models lead to more accurate predictions, and leverage user input much more effectively than state-of-the-art independent models.

6.1.5. Metric Learning for Large Scale Image Classification: Generalizing to new classes at near-zero cost

Participants: Gabriela Csurka [XRCE], Thomas Mensink, Florent Perronnin [XRCE], Jakob Verbeek.

In [18], [27] we consider the task of large scale image classification in open ended datasets. Many real-life datasets are open-ended and dynamic: new images are continuously added to existing classes, new classes appear over time and the semantics of existing classes might evolve too. In order to be able to handle new images and new classes at near-zero cost we consider two distance based classifiers, the k-nearest neighbor (k-NN) and nearest class mean (NCM) classifiers. For the NCM classifier we introduce a new metric learning approach, which has advantageous properties over the classical Fisher Discriminant Analysis. We also introduce an extension of the NCM classifier to allow for richer class representations, using multiple centroids per class. Experiments on the ImageNet 2010 challenge dataset, which contains over one million training images of thousand classes, show that, surprisingly, the NCM classifier compares favorably to the more flexible k-NN classifier. Moreover, the NCM performance is comparable to that of linear SVMs which obtain current state-of-the-art performance. Experimentally we study the generalization performance to classes that were not used to learn the metrics. Using a metric learned on 1,000 classes, we show results for the ImageNet-10K dataset which contains 10,000 classes, and obtain performance that is competitive with the current state-of-the-art, while being orders of magnitude faster. Furthermore, we show how a zero-shot class prior based on the ImageNet hierarchy can improve performance when few training images are available. See Figure 4 for an illustration.

6.2. Learning and statistical models

6.2.1. Image categorization using Fisher kernels of non-iid image models

Participants: Ramazan Cinbis, Cordelia Schmid, Jakob Verbeek.

Bag of visual words treat images as an orderless sets of local regions and represent them by visual word frequency histograms. Implicitly, regions are assumed to be identically and independently distributed (iid), which is a very poor assumption from a modeling perspective; see Figure 5 for an illustration. In [13], we introduce non-iid models by treating the parameters of bag-of-word models as latent variables which are integrated out, rendering all local regions dependent. Using the Fisher kernel we encode an image by the gradient of the data log-likelihood with respect to hyper-parameters that control priors on the model parameters. In fact, our models naturally generate transformations similar to taking square-roots, providing an explanation of why such non-linear transformations have proven successful. Using variational inference we extend the basic model to include Gaussian mixtures over local descriptors, and latent topic models to capture the co-occurrence structure of visual words, both improving performance. Our models yields state-of-the-art image categorization performance using linear classifiers, without using non-linear kernels, or (approximate) explicit embeddings thereof, e.g. by taking the square-root of the features.

6.2.2. Conditional gradient algorithms for machine learning

Participants: Zaid Harchaoui, Anatoli Juditsky [UJF], Arkadi Nemirovski [Georgia Tech].

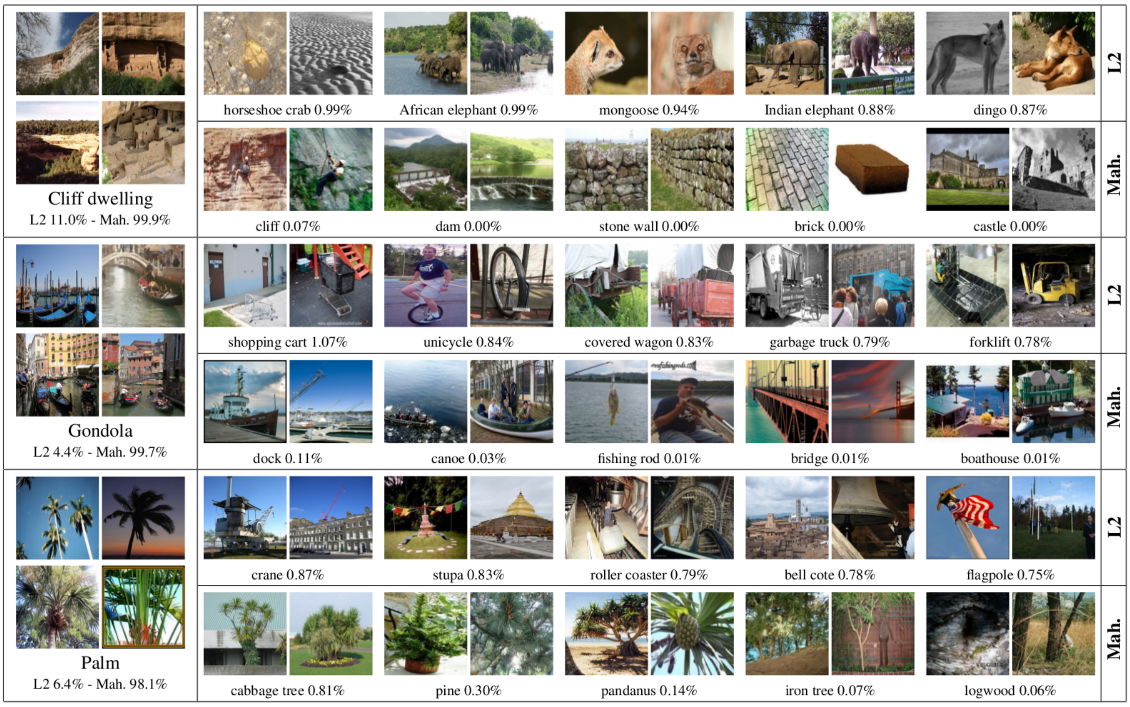


Figure 4. Examples of three classes, and the five most similar classes for each according to the standard ℓ_2 metric and our learned Mahalanobis metric.

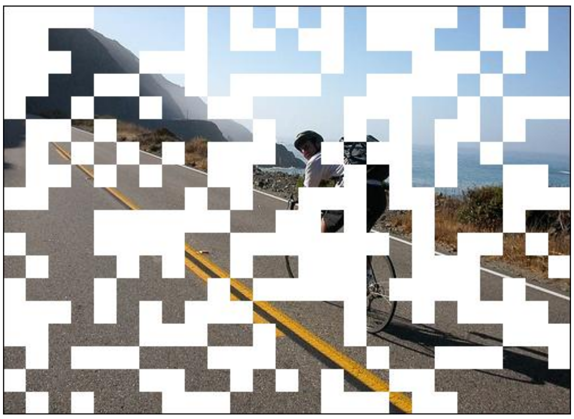


Figure 5. Illustration of why local image patches are not independent: we can easily guess the image content in the masked areas.

In [17] we consider convex optimization problems arising in machine learning in high-dimensional settings. For several important learning problems, such as e.g. noisy matrix completion, state-of-the-art optimization approaches such as composite minimization algorithms are difficult to apply and do not scale up to large datasets. We study three conditional gradient-type algorithms, suitable for large-scale problems, and derive their finite-time convergence guarantees. Promising experimental results are presented on two large-scale real-world datasets.

6.2.3. Large-scale classification with trace-norm regularization

Participants: Matthijs Douze, Miro Dudik [Microsoft Research], Zaid Harchaoui, Jérôme Malick [BiPoP Team Inria Grenoble], Mattis Paulin [ETHZ].

In [16] we introduce a new scalable learning algorithm for large-scale multi-class image classification, based on the multinomial logistic loss and the trace-norm regularization penalty. Reframing the challenging non-smooth optimization problem into a surrogate infinite-dimensional optimization problem with a regular ℓ_1-regularization penalty, we propose a simple and provably efficient accelerated coordinate descent algorithm. Furthermore, we show how to perform efficient matrix computations in the compressed domain for quantized dense visual features, scaling up to 100,000s examples, 1,000s-dimensional features, and 100s of categories. Promising experimental results on the "Fungus", "Ungulate", and "Vehicles" subsets of ImageNet are presented, where we show that our approach performs significantly better than state-of-the-art approaches for Fisher vectors with 16 Gaussians.

6.2.4. Tree-walk kernels for computer vision

Participants: Francis Bach [Inria SIERRA team], Zaid Harchaoui.

In [25] we propose a family of positive-definite kernels between images, allowing to compute image similarity measures respectively in terms of color and of shape. The kernels consists in matching subtree-patterns called "tree-walks" of graphs extracted from the images, e.g. the segmentation graphs for color similarity and graphs of the discretized shapes or the point clouds in general for shape similarity. In both cases, we are able to design computationally efficient kernels which can be computed in polynomial-time in the size of the graphs, by leveraging specific properties of the graphs at hand such as planarity for segmentation graphs or factorizability of the associated graphical model for point clouds. Our kernels can be used by any kernel-based learning method, and hence we present experimental results for supervised and semi-supervised classification as well as clustering of natural images and supervised classification of handwritten digits and Chinese characters from few training examples.

6.2.5. Lifted coordinate descent for learning with trace-norm regularization

Participants: Miro Dudik [Microsoft Research], Zaid Harchaoui, Jérôme Malick [BiPoP Team Inria Grenoble].

In [14] we consider the minimization of a smooth loss with trace-norm regularization, which is a natural objective in multi-class and multi-task learning. Even though the problem is convex, existing approaches rely on optimizing a non-convex variational bound, which is not guaranteed to converge, or repeatedly perform singular-value decomposition, which prevents scaling beyond moderate matrix sizes. We lift the non-smooth convex problem into an infinitely dimensional smooth problem and apply coordinate descent to solve it. We prove that our approach converges to the optimum, and is competitive or outperforms the state of the art.

6.3. Recognition in video

6.3.1. Large-scale multi-media event detection in video

Participants: Matthijs Douze, Zaid Harchaoui, Dan Oneata, Danila Potapov, Jerome Revaud, Cordelia Schmid, Jochen Schwenninger [Fraunhofer Institute, Bonn], Jakob Verbeek, Heng Wang.

This year we participated in the TrecVid Multimedia Event Detection (MED) task. The goal is to detect events categories (such as "birthday party", or "changing a vehicle tire") in a large collection of around 100,000 videos with a total duration of around 4,000 hours. To this end we implemented an efficient system based on our recently developed MBH video descriptor (see Section 5.4) , SIFT descriptors and, MFCC audio descriptors (contributed by Fraunhofer Institute). All these low-level descriptors are encoded using the Fisher vector representation (see Section 5.3). In addition we implemented an optical character recognition (OCR) system to extract textual features from the video. The system is described in a forthcoming paper [31], and ranked first and second in two evaluations among the 17 systems submitted by different international teams participating to the task. See Figure 6 for an illustration.

Figure 6. Illustration of videos retrieved for two event categories. From left to right, we show for each a frame from (i) the top ranked video, (ii,iii) the first negative video, and the postive just before, and (iv) the last positive video.

6.3.2. Learning Object Class Detectors from Weakly Annotated Video

Participants: Javier Civera, Vittorio Ferrari, Christian Leistner, Alessandro Prest, Cordelia Schmid.

Object detectors are typically trained on a large set of still images annotated by bounding-boxes. In [20] we introduce an approach for learning object detectors from real-world web videos known only to contain objects of a target class. We propose a fully automatic pipeline that localizes objects in a set of videos of the class and learns a detector for it. The approach extracts candidate spatio-temporal tubes based on motion segmentation and then selects one tube per video jointly over all videos. See Figure 7 for an illustration. To compare to the state of the art, we test our detector on still images, i.e., Pascal VOC 2007. We observe that frames extracted from web videos can differ significantly in terms of quality to still images taken by a good camera. Thus, we formulate the learning from videos as a domain adaptation task. We show that training from a combination of weakly annotated videos and fully annotated still images using domain adaptation improves the performance of a detector trained from still images alone.

6.3.3. Recognizing activities with cluster-trees of tracklets

Participants: Adrien Gaidon, Zaid Harchaoui, Cordelia Schmid.

In [15] we address the problem of recognizing complex activities, such as pole vaulting, which are characterized by the composition of a large and variable number of different spatio-temporal parts. We represent a video as a hierarchy of mid-level motion components. This hierarchy is a data-driven decomposition specific to each video. We introduce a divisive clustering algorithm that can efficiently extract a hierarchy over a large

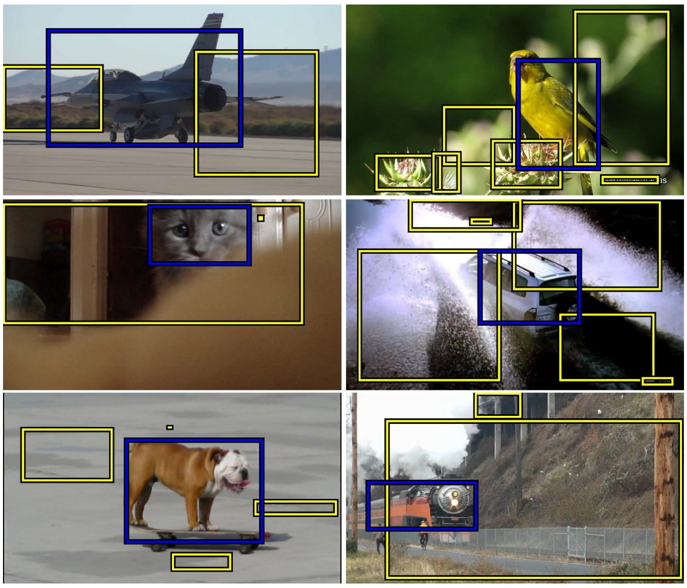


Figure 7. Yellow boxes represent tubes extracted by our method on the YouTube-Objects dataset. Blue boxes indicate the automatically selected tubes.

number of local trajectories. We use this structure to represent a video as an unordered binary tree. This tree is modeled by nested histograms of local motion features, see Figure 8 . We provide an efficient positive definite kernel that computes the structural and visual similarity of two tree decompositions by relying on models of their edges. Contrary to most approaches based on action decompositions, we propose to use the full hierarchical action structure instead of selecting a small fixed number of parts. We present experimental results on two recent challenging benchmarks that focus on complex activities and show that our kernel on per-video hierarchies allows to efficiently discriminate between complex activities sharing common action parts. Our approach improves over the state of the art, including unstructured activity models, baselines using other motion decomposition algorithms, graph matching, and latent models explicitly selecting a fixed number of parts.

6.3.4. Action Detection with Actom Sequence Models

Participants: Adrien Gaidon, Zaid Harchaoui, Cordelia Schmid.

We address the problem of detecting actions, such as drinking or opening a door, in hours of challenging video data. In [26] we propose a model based on a sequence of atomic action units, termed "actoms", that are semantically meaningful and characteristic for the action. Our Actom Sequence Model (ASM) represents the temporal structure of actions as a sequence of histograms of actom-anchored visual features, see Figure 9 for an illutration. Our representation, which can be seen as a temporally structured extension of the bag-of-features, is flexible, sparse, and discriminative. Training requires the annotation of actoms for action examples. At test time, actoms are detected automatically based on a non-parametric model of the distribution of actoms, which also acts as a prior on an action's temporal structure. We present experimental results on two recent benchmarks for temporal action detection: "Coffee and Cigarettes" and the "DLSB" dataset. We also adapt our approach to a classification by detection set-up and demonstrate its applicability on the challenging "Hollywood 2" dataset. We show that our ASM method outperforms the current state of the art in temporal action detection, as well as baselines that detect actions with a sliding window method combined with bag-of-features.

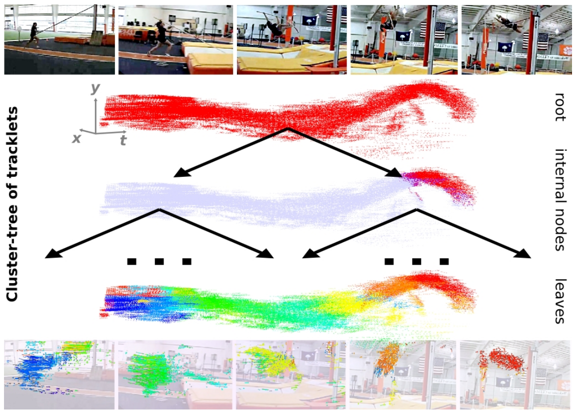


Figure 8. Illustration of tracklets found in a video and their hierarchical decomposition.

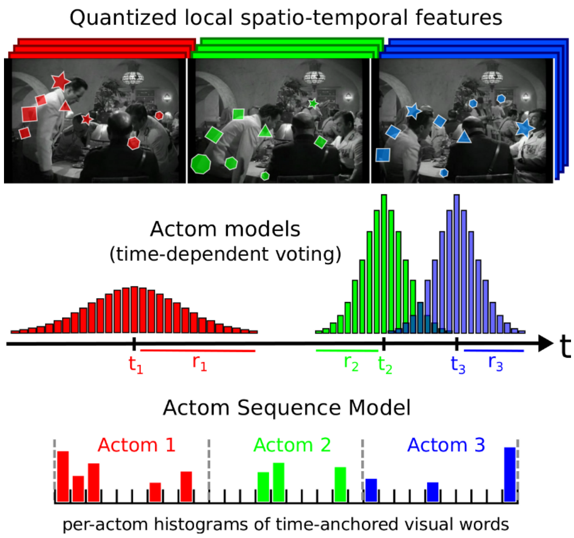


Figure 9. Illustration of the "Actom" video representation, see text for details.

6.3.5. Action recognition by dense trajectories

Participants: Alexander Kläser, Cheng-Lin Liu [Chinese Academy of Sciences], Cordelia Schmid, Heng Wang [Chinese Academy of Sciences].

In [28] we introduce a video representation based on dense trajectories and motion boundary descriptors. Trajectories capture the local motion information of the video. A state-of-the-art optical flow algorithm enables a robust and efficient extraction of the dense trajectories. As descriptors we extract features aligned with the trajectories to characterize shape (point coordinates), appearance (histograms of oriented gradients) and motion (histograms of optical flow). Additionally, we introduce a descriptor based on motion boundary histograms (MBH) (see the visualization in Figure 10), which is shown to consistently outperform other state-of-the-art descriptors, in particular on real-world videos that contain a significant amount of camera motion. We evaluate our video representation in the context of action classification on nine datasets, namely KTH, YouTube,Hollywood2, UCF sports, IXMAS, UIUC, Olympic Sports, UCF50 and HMDB51. On all datasets our approach outperforms current state-of-the-art results.

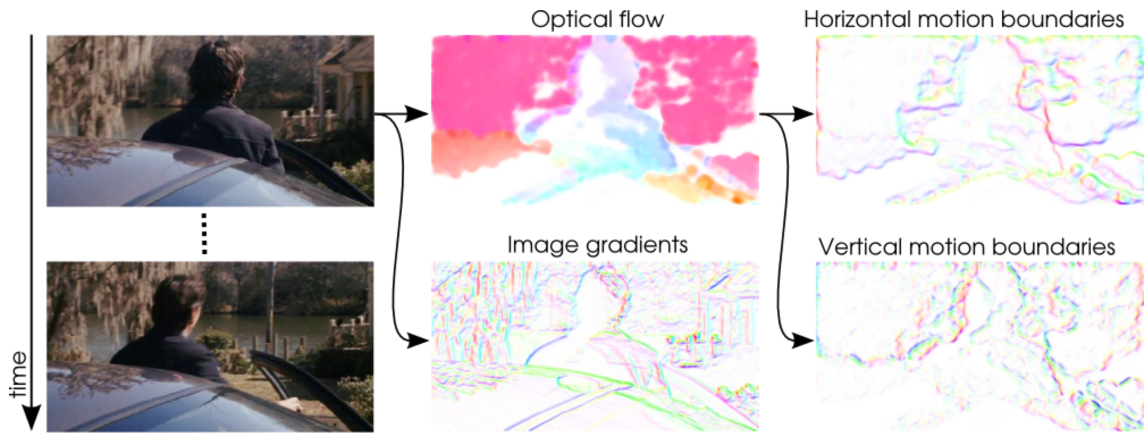


Figure 10. Illustration of the information captured by HOG, HOF, and MBH descriptors. Gradient/flow orientation is indicated by color (hue) and magnitude by saturation. The optical flow (top, middle) shows constant motion in the background, which is due to the camera movements. The motion boundaries (right) encode the relative motion between the person and the background.

MAGRIT Project-Team

6. New Results

6.1. Motion, scene and camera reconstruction

Participants: Marie-Odile Berger, Srikrishna Bhat, Christel Leonet, Gilles Simon, Frédéric Sur.

- **Enhancing the grid method for in-plane strain measurements**
  This work is motivated by a problem from experimental solid mechanics. The grid method permits to measure the heterogeneous strains on the surface of specimens subjected to mechanical tests. Among full-field measurement techniques, the grid method consists in transferring a regular grid on the surface of the specimen and in taking images of the grid before and after deformation. Windowed Fourier analysis then gives an estimate of the surface displacement and strain components. In a collaboration with Institut Pascal (Université Blaise Pascal, Clermont Ferrand), we have shown that the estimations obtained by this technique are approximately the convolution of the actual values with the analysis window. We have also characterized how the noise in the grid image impairs the displacement and strain maps [18]. This study has allowed us to improve the metrological performance of the grid method with deconvolution algorithms. A numerical and experimental study can be found in [17].
- **Visual words for pose computation**
  Visual vocabularies are standard tools in the object/image classification literature, and are emerging as a new tool for building point correspondences for pose estimation. Within S. Bhat's PhD thesis, we have proposed several methods for visual word construction dedicated to point matching, with structure from motion and pose estimation applications in view. The three dimensional geometry of a scene is first extracted with bundle adjustment techniques based on keypoint correspondences. These correspondences are obtained by grouping the set of all SIFT descriptors from the training images into visual words using transitive closure (TC) techniques. We obtain a more accurate 3D geometry than with classical image-to-image point matching. In a second on-line step, these visual words serve as 3D point descriptors that are robust to viewpoint change, and are used for building 2D-3D correspondences on-line during application, yielding the pose of the camera by solving the PnP problem. Several visual word formation techniques have been compared with respect to robustness to viewpoint change between the learning and the test images. Our experiments showed that the adaptive TC visual words are better in many ways when compared to other classical techniques such as K-means.

  More specifically, the work of this year has focused on improving pose estimation from visual words with respect to strong viewpoint changes. 2D-3D correspondences are actually difficult to establish if there are too large viewpoint changes between the image whose pose is sought and the images that yielded the visual words attached to 3D points. We assessed several viewpoint simulation techniques in order to enrich the visual word description of the 3D points.
- **Acquisition of 3D calibrated data**
  Christel Leonet joined the team in October 2010 as an Inria assistant engineer with the aim of building an integrated 3D acquisition system. More specifically, the objective of her work is to combine an IMU (Inertial Measurement Unit), a GPS receiver, a laser rangefinder and a video camera for ground truth data acquisitions of camera movements and scene structures. These data will be useful to validate several algorithms developed in our team. This year, a new visual pan tracking method has been designed and implemented. We considered spherical environments made of sparse video images instead of fully-covered environment maps which often suffer from geometric and photometric misalignments. The scanning process has been improved in order to increase the accuracy of the recovered polygons and allow for visual assessments of this accuracy. The 3D laser pointer has been validated in several indoor environments. Finally, the GPS has been integrated to the system and preliminary results have been obtained in outdoor environments.

6.2. Medical imaging

Participants: René Anxionnat, Marie-Odile Berger, Nazim Haouchine, Erwan Kerrien, Pierre-Frédéric Villard, Brigitte Wrobel-Dautcourt, Ahmed Yureidini.

- **Vessel reconstruction with implicit surfaces**
  This research activity is led in collaboration with Shacra project-team from Inria Lille-Nord Europe and the Department of Interventional Neuroradiology from Nancy University Hospital. It was pursued this year in the context of the SOFA-InterMedS Inria Large-Scale Initiative (http://www.sofa-framework.org/).

  Our objective is to offer the interventional radiologists with a patient-based interactive simulator [16]. The medical applications are training to endovascular procedures, planning the intervention, and augmenting the intra-operative images with 3D simulated data. Our contributions address vasculature modeling from patient data, namely 3D rotational angiography (3DRA) volumes. The segmentation should be both user friendly and generate a vascular surface model that is compliant with the computing constraints set in interactive simulation. Within A. Yureidini's PhD thesis, a new model was developed consisting of a tree of local implicit blobby models. The algorithm consists of two steps: first, a vessel tracking step to extract the vessel topology and, second, fitting local surface data points with implicit blobby models at each node point on the vessel centerline.

  An extensive validation of our RANSAC-based vessel tracking algorithm was performed [14], by comparison with state of the art Multiple Hypothesis Testing [19] on 10 patient data. Fitting the implicit model to patient data relies on the minimization of a multi-termed energy. A closed form solution was derived, and a blob selection and subdivision heuristic was described to implement an efficient energy minimization algorithm. Both the geometric accuracy and compactness of the resulting vascular models were shown to be excellent [15].

  Our current goals are: first, to further enhance model compactness by relying on the robustness and versatility of the modeling algorithm and using sparser vascular centerline trees; second, to mathematically ensure the continuity between neighboring local implicit models; and third, to reintroduce the raw image data for a more accurate energy computation, with the aim to design a blobby deformable model.

  This model was implemented in Sofa simulation platform, enabling interactive simulation time and thereby showing an impressive realism during tool navigation. On-going preliminary medical evaluation is being carried on by our fellow interventional radiologist in the framework of intervention planning.

- **Designing respiration models for patient based simulators**
  The work presented here has been done within a collaboration with Imperial College of London, Bangor University and Inria Aviz team.

  Respiratory models could be a key component in increasing realism in medical simulators. We have previously developed such kind of model. However finding the good parameters to tune the model so that it corresponds to a real patient behavior is not an easy task.

  This year, we have studied methods to automatically tune the elasticity of soft-tissues and the respiratory model parameters based on patient data. The estimation is based on two 3D Computed Tomography scans of the same patient at two different time steps. The parametrization of the model is considered as an inverse problem. Optimization techniques have then been deployed to solve the problem.

  In [13], we used a random search algorithm to generate a given number of sets of 15 random parameters. The set of parameters that provides the lowest fitness is extracted and corresponds to the solution of the optimization problem.

In [9], we have made use of an ad-hoc evolutionary algorithm that is able to explore a search space with 15 dimensions. Our method is fully automatic and auto-adaptive. A compound fitness function has been designed to account for various quantities that have to be minimized. The algorithm efficiency was experimentally analyzed on several real test-cases: i) three patient datasets have been acquired with the "breath hold" protocol, and ii) two datasets corresponds to 4D CT scans. The performance was compared with two traditional methods (downhill simplex and conjugate gradient descent), our random search method and a basic real-valued genetic algorithm. The results showed that our evolutionary scheme provides more significantly stable and accurate results.

- **Physics-based augmented reality**
  The development of AR systems for use in the medical field faces one major challenge: the correct superposition of pre-operative data onto intraoperative images. This task is especially difficult when laparospic surgery is considered since superposition must be achieved on deformable organs. Most existing AR systems only consider rigid registration between the pre and intraoperative data and the transformation is often computed interactively or from markers attached to the patient's body. In cooperation with the Shacra team, we have introduced an original method to perform augmented or mixed reality on deformable objects. Compared to state-of-the-art techniques, our method is able to track deformations of volumetric objects and not only surfacic objects. A flexible framework that relies on the combination of 3D motion estimation obtained from stereoscopic data and a physics-based deformable model used as a regularization and interpolation step allows us to perform non-rigid and robust registration between the pre and intraoperative images [10].

MAIA Project-Team

6. New Results

6.1. Decision Making

6.1.1. Accounting for Uncertainty in Penetration Testing

Participants: Olivier Buffet, Jörg Hoffmann.

Carlos Sarraute (Core Security Technologies) is an external collaborator.

Core Security Technologies is an U.S.-American/Argentinian company providing, amongst other things, tools for (semi-)automated security checking of computer networks against outside hacking attacks. For automation of such checks, a module is needed that automatically generates potential attack paths. Since the application domain is highly dynamic, a module allowing to declaratively specify the environment (the network and its configuration) is highly advantageous. For that reason, Core Security Technologies have been looking into using AI Planning techniques for this purpose. After consulting by Jörg Hoffmann, they are now using a variant of Jörg Hoffmann's FF planner in their product. While that solution is satisfactory in many respects, it also has weaknesses. The main weakness is that it does not handle the incomplete knowledge in this domain – figuratively speaking, the attacker is assumed to have perfect information about the network. This results in high costs in terms of runtime and network traffic, for extensive scanning activities prior to planning.

We are currently working with Core Security's research department to overcome this issue, by modeling and solving the attack planning problem as a POMDP instead. A workshop paper detailing the POMDP model has been published at SecArt'11. While such a model yields much higher quality attacks, solving an entire network as a POMDP is not feasible. We have designed a decomposition method making use of network structure and approximations to overcome this problem, by using the POMDP model only to find good-quality attacks on single machines, and propagating the results through the network in an appropriate manner. This work has been published in ICAPS'12 [34].

6.1.2. Searching for Information with MDPs

Participants: Mauricio Araya, Olivier Buffet, Vincent Thomas, François Charpillet.

In the context of Mauricio Araya's PhD, we are working on how MDPs —or related models— can search for information. This has led to various research directions, such as extending POMDPs so as to optimize information-based rewards, or actively learning MDP models. This year, we have focused on a novel optimistic Bayesian Reinforcement Learning algorithm –as described below– and on Mauricio's dissertation.

Exact or approximate solutions to Model-based Bayesian RL are impractical, so that a number of heuristic approaches have been considered, most of them relying on the principle of "optimism in the face of uncertainty". Some of these algorithms have properties that guarantee the quality of their outcome, inspired by the PAC-learning (Probably Approximately Correct) framework. For example, some algorithms provably make in most cases the same decision as would be made if the true model were known (PAC-MDP property).

We have proposed a novel optimistic algorithm, BOLT, that is

- appealing in that it is (i) optimistic *about* the uncertainty in the model and (ii) deterministic (thus easier to study); and
- provably PAC-BAMDP, i.e., makes in most cases the same decision as a perfect BRL algorithm would.

This work has been published in ICML'12 [9] and (in French) in JFPDA'12 [30], additional details appearing in [40].

6.1.3. Scheduling for Probabilistic Realtime Systems

Participant: Olivier Buffet.

Maxim Dorin, Luca Santinelli, Liliana Cucu-Grosjean (Inria, TRIO team), and Rob Davies (U. of York) are external collaborators.

In this collaborative research work (mainly with the TRIO team), we look at the problem of scheduling periodic tasks on a single processor, in the case where each task's period is a (known) random variable. In this setting, some job will necessarily be missed, so that one will try to satisfy some criteria depending on the number of deadline misses.

We have proposed three criteria: (1) satisfying pre-defined deadline miss ratios, (2) minimizing the worst deadline miss ratio, and (3) minimizing the average deadline miss ratio. For each criterion we propose an algorithm that computes a provably optimal fixed priority assignment, i.e., a solution obtained by assigning priorities to tasks and executing jobs by order of priority.

This work has been presented in RTNS'11, and an extended version is currently in preparation.

6.1.4. Adaptive Management with POMDPs

Participant: Olivier Buffet.

Iadine Chadès, Josie Carwardine, Tara G. Martin (CSIRO), Samuel Nicol (U. of Alaska Fairbanks) and Régis Sabbadin (INRA) are external collaborators.

In the field of conservation biology, adaptive management is about managing a system, e.g., performing actions so as to protect some endangered species, while learning how it behaves. This is a typical reinforcement learning task that could for example be addressed through BRL.

Here, we consider that a number of experts provide us with one possible model each, assuming that one of them is the true model. This allows making decisions by solving a *hidden model MDP* (hmMDP). An hmMDP is essentially a simplified mixed observability MDP (MOMDP), where the hidden part of the state corresponds to the model (in cases where all other variables are fully observable).

From a theoretical point of view, we have proved that deciding whether a finite-horizon hmMDP problem admits a solution policy of value greater than a pre-defined threshold is a PSPACE-complete problem. We have also conducted preliminary studies of this approach, using the scenario of the protection of the Gouldian finch, and focusing on the particular characteristics that could be exploited to more efficiently solve this problem. These results have been presented in AAAI'12 [14].

6.1.5. Multi-Camera Tracking in Partially Observable Environment

Participants: Arsène Fansi Tchango, Olivier Buffet, Vincent Thomas, Alain Dutech.

Fabien Flacher (Thales THERESIS) is an external collaborator.

In collaboration with Thales ThereSIS - SE&SIM Team (Synthetic Environment & Simulation), we focus on the problem of following the trajectories of several persons with the help of several actionable cameras. This problem is difficult since the set of cameras cannot cover simultaneously the whole environment, since some persons can be hidden by obstacles or by other persons, and since the behavior of each person is governed by internal variables which can only be inferred (such as his motivation or his hunger).

The approach we are working on is based on (1) POMDP formalisms to represent the state of the system (person and their internal states) and possible actions for the cameras, (2) a simulator provided and developed by Thales ThereSIS and (3) particle filtering approaches based on this simulator.

From a theoretical point of view, we are currently investigating how to use a deterministic simulator and to generate new particles in order to keep a good approximation of the posterior distribution.

6.1.6. Scaling Up Decentralized MDPs Through Heuristic Search

Participant: Jilles Dibangoye.

External collaborators: Christopher Amato, Arnaud Doniec.

Decentralized partially observable Markov decision processes (Dec-POMDPs) are rich models for cooperative decision-making under uncertainty, but are often intractable to solve optimally (NEXP-complete). The transition and observation independent Dec-MDP is a general subclass that has been shown to have complexity in NP, but optimal algorithms for this subclass are still inefficient in practice. We first provide an updated proof that an optimal policy does not depend on the histories of the agents, but only the local observations. We then present a new algorithm based on heuristic search that is able to expand search nodes by using constraint optimization. We show experimental results comparing our approach with the state-of-the-art Dec-MDP and Dec-POMDP solvers. These results show a reduction in computation time and an increase in scalability by multiple orders of magnitude in a number of benchmarks.

This work was presented in UAI'2012 [16].

6.1.7. Approximate Modified Policy Iteration

Participant: Bruno Scherrer.

External collaborators: Victor Gabillon, Mohammad Ghavamzadeh and Matthieu Geist.

Modified policy iteration (MPI) is a dynamic programming (DP) algorithm that contains the two celebrated policy and value iteration methods. Despite its generality, MPI has not been thoroughly studied, especially its approximation form which is used when the state and/or action spaces are large or infinite. We have proposed three implementations of approximate MPI (AMPI) that are extensions of well-known approximate DP algorithms: fitted-value iteration, fitted-Q iteration, and classification-based policy iteration. We have provided an error propagation analysis that unifies those for approximate policy and value iteration. For the classification-based implementation, we have developed a finite-sample analysis that shows that MPI's main parameter allows to control the balance between the estimation error of the classifier and the overall value function approximation.

This work was presented in JFPDA'2012 [35] and ICML'2012 [45].

6.1.8. A Dantzig Selector Approach to Temporal Difference Learning

Participant: Bruno Scherrer.

External collaborators: Matthieu Geist, Mohammad Ghavamzadeh and Alessandro Lazaric.

LSTD is one of the most popular reinforcement learning algorithms for value function approximation. Whenever the number of samples is larger than the number of features, LSTD must be paired with some form of regularization. In particular, L_1-regularization methods tend to perform feature selection by promoting sparsity and thus they are particularly suited in high-dimensional problems. Nonetheless, since LSTD is not a simple regression algorithm but it solves a fixed-point problem, the integration with L_1-regularization is not straightforward and it might come with some drawbacks (see e.g., the P-matrix assumption for LASSO-TD). We have introduced a novel algorithm obtained by integrating LSTD with the Dantzig Selector. In particular, we have investigated the performance of the algorithm and its relationship with existing regularized approaches, showing how it overcomes some of the drawbacks of existing solutions.

This work was presented at JFPDA'2012 [33] and ICML'2012 [20].

6.1.9. On the Use of Non-Stationary Policies for Stationary Infinite-Horizon Markov Decision Processes

Participants: Bruno Scherrer, Boris Lesner.

In infinite-horizon stationary γ-discounted Markov Decision Processes, it is known that there exists a stationary optimal policy. Using Value and Policy Iteration with some error ϵ at each iteration, it is well-known that one can compute stationary policies that are $\frac{2\gamma}{(1-\gamma)^2}\epsilon$-optimal. After having shown that this guarantee is tight, we have developed variations of Value and Policy Iteration for computing non-stationary policies that can be up to $\frac{2\gamma}{1-\gamma}\epsilon$-optimal, which constitutes a significant improvement in the usual situation when γ is close to 1. Surprisingly, this shows that the problem of "computing near-optimal non-stationary policies" is much simpler than that of "computing near-optimal stationary policies".

This work was presented and selected for a full oral presentation at NIPS'2012 [28].

6.1.10. Developmental Reinforcement Learning

Participant: Alain Dutech.

External collaborators: Matthieu Geist (IMS Supelec), Olivier Pietquin (IMS Supelec)

Reinforcement Learning in rich, complex and large sensorimotor spaces is a difficult problem mainly because the exploration of such a huge space cannot be done in an extensive way. The idea is thus to adopt a developmental approach where the perception and motor skills of the robot can grow in richness and complexity during learning, as a consequence the size of the state and action spaces grows progressively when the performances of the learning agent increases. The learning framework relies on function approximators with specific properties (continuous input space, life-long adaptation, knowledge transfer). Architectures based on "reservoir learning" and "dynamical self-organizing maps" kind of artificial neural networks have been investigated [32], [18].

6.1.11. Dialog and POMDPs

Participant: Lucie Daubigney.

Reinforcement learning (RL) is now part of the state of the art in the domain of spoken dialog systems (SDS) optimization. The best performing RL methods, such as those based on Gaussian Processes, require to test small changes in the policy to assess them as improvements or degradations. This process is called on policy learning. Nevertheless, it can result in system behaviors that are not acceptable by users. Learning algorithms should ideally infer an optimal strategy by observing interactions generated by a non-optimal but acceptable strategy, that is learning off-policy. Such methods usually fail to scale up and are thus not suited for real-world systems. In this work, a sample-efficient, on-line and off-policy RL algorithm is proposed to learn an optimal policy [15]. This algorithm is combined to a compact non-linear value function representation (namely a multilayer perceptron) enabling to handle large scale systems. One of the application domain is the teaching of a second language [31].

6.1.12. SAP Speaks PDDL: Exploiting a Software-Engineering Model for Planning in Business Process Management

Participant: Jörg Hoffmann.

Ingo Weber (NICTA) and Frank Michael Kraft (bpmnforum.net) are external collaborators.

Planning is concerned with the automated solution of action sequencing problems described in declarative languages giving the action preconditions and effects. One important application area for such technology is the creation of new processes in Business Process Management (BPM), which is essential in an ever more dynamic business environment. A major obstacle for the application of Planning in this area lies in the modeling. Obtaining a suitable model to plan with – ideally a description in PDDL, the most commonly used planning language – is often prohibitively complicated and/or costly. Our core observation in this work is that this problem can be ameliorated by leveraging synergies with model-based software development. Our application at SAP, one of the leading vendors of enterprise software, demonstrates that even one-to-one model re-use is possible.

The model in question is called Status and Action Management (SAM). It describes the behavior of Business Objects (BO), i.e., large-scale data structures, at a level of abstraction corresponding to the language of business experts. SAM covers more than 400 kinds of BOs, each of which is described in terms of a set of status variables and how their values are required for, and affected by, processing steps (actions) that are atomic from a business perspective. SAM was developed by SAP as part of a major model-based software engineering effort. We show herein that one can use this same model for planning, thus obtaining a BPM planning application that incurs no modeling overhead at all.

We compile SAM into a variant of PDDL, and adapt an off-the-shelf planner to solve this kind of problem. Thanks to the resulting technology, business experts may create new processes simply by specifying the desired behavior in terms of status variable value changes: effectively, by describing the process in their own language.

This work has been published in JAIR [6].

6.1.13. Resource-Constrained Planning: A Monte Carlo Random Walk Approach

Participant: Jörg Hoffmann.

Hootan Nakhost and Martin Müller (University of Alberta) are external collaborators.

The need to economize limited resources, such as fuel or money, is a ubiquitous feature of planning problems. If the resources cannot be replenished, the planner must make do with the initial supply. It is then of paramount importance how constrained the problem is, i.e., whether and to which extent the initial resource supply exceeds the minimum need. While there is a large body of literature on numeric planning and planning with resources, such resource constrainedness has only been scantily investigated. We herein start to address this in more detail. We generalize the previous notion of resource constrainedness, characterized through a numeric problem feature $C \leq 1$, to the case of multiple resources. We implement an extended benchmark suite controlling C. We conduct a large-scale study of the current state of the art as a function of C, highlighting which techniques contribute to success. We introduce two new techniques on top of a recent Monte Carlo Random Walk method, resulting in a planner that, in these benchmarks, outperforms previous planners when resources are scarce (C close to 1). We investigate the parameters influencing the performance of that planner, and we show that one of the two new techniques works well also on the regular IPC benchmarks.

This work has been published in ICAPS-12 [26].

6.1.14. How to Relax a Bisimulation?

Participants: Michael Katz, Jörg Hoffmann.

Malte Helmert (Basel University) is an external collaborator.

Merge-and-shrink abstraction (M&S) is an approach for constructing admissible heuristic functions for cost-optimal planning. It enables the targeted design of abstractions, by allowing to choose individual pairs of (abstract) states to aggregate into one. A key question is how to actually make these choices, so as to obtain an informed heuristic at reasonable computational cost. Recent work has addressed this via the well-known notion of bisimulation. When aggregating only bisimilar states – essentially, states whose behavior is identical under every planning operator – M&S yields a perfect heuristic. However, bisimulations are typically exponentially large. Thus we must relax the bisimulation criterion, so that it applies to more state pairs, and yields smaller abstractions. We herein devise a fine-grained method for doing so. We restrict the bisimulation criterion to consider only a subset K of the planning operators. We show that, if K is chosen appropriately, then M&S still yields a perfect heuristic, while abstraction size may decrease exponentially. Designing practical approximations for K, we obtain M&S heuristics that are competitive with the state of the art.

This work has been published in ICAPS-12 [22], and as Inria research report RR-7901 [42].

6.1.15. Semi-Relaxed Plan Heuristics

Participants: Emil Keider, Jörg Hoffmann.

Patrik Haslum (ANU) is an external collaborator.

Heuristics based on the delete relaxation are at the forefront of modern domain-independent planning techniques. Here we introduce a principled and flexible technique for augmenting delete-relaxed tasks with a limited amount of delete information, by introducing special fluents that explicitly represent conjunctions of fluents in the original planning task. Differently from previous work in this direction, conditional effects are used to limit the growth of the task to be linear, rather than exponential, in the number of conjunctions that are introduced, making its use for obtaining heuristic functions feasible. We discuss how to obtain an informative set of conjunctions to be represented explicitly, and analyze and extend existing methods for relaxed planning in the presence of conditional effects. The resulting heuristics are empirically evaluated, and shown to be sometimes much more informative than standard delete-relaxation heuristics.

This work has been published in ICAPS-12 [24].

6.1.16. Structural Patterns Beyond Forks: Extending the Complexity Boundaries of Classical Planning

Participants: Michael Katz, Emil Keider.

Tractability analysis in terms of the causal graphs of planning problems has emerged as an important area of research in recent years, leading to new methods for the derivation of domain-independent heuristics (Katz and Domshlak 2010). Here we continue this work, extending our knowledge of the frontier between tractable and NP-complete fragments. We close some gaps left in previous work, and introduce novel causal graph fragments that we call the hourglass and semi-fork, for which under certain additional assumptions optimal planning is in P. We show that relaxing any one of the restrictions required for this tractability leads to NP-complete problems. Our results are of both theoretical and practical interest, as these fragments can be used in existing frameworks to derive new abstraction heuristics. Before they can be used, however, a number of practical issues must be addressed. We discuss these issues and propose some solutions.

This work has been published in AAAI-12 [23].

6.2. Understanding and mastering complex systems

6.2.1. Adaptive control of a complex system based on its multi-agent model

Participants: Vincent Chevrier, Tomas Navarrete.

Laurent Ciarletta (Madynes team, LORIA) is an external collaborator.

Complex systems are present everywhere in our environment: internet, electricity distribution networks, transport networks. These systems have the following characteristics: a large number of autonomous entities, dynamic structures, different time and space scales and emergent phenomena. This work is centered on the problem of control of such systems. The problem is defined as the need to determine, based on a partial perception of the system state, which actions to execute in order to avoid or favor certain global states of the system. This problem comprises several difficult questions: how to evaluate the impact at the global level of actions applied at a global level, how to model the dynamics of an heterogeneous system (different behaviors issue of different levels of interactions), how to evaluate the quality of the estimations issue of the modeling of the system dynamics.

We propose a control architecture[1] based on an “equation-free” approach. We use a multi-agent model to evaluate the global impact of local control actions before applying the most pertinent set of actions.

Associated to our architecture, an experimental platform has been developed to confront the basic ideas or the architecture within the context of simulated “free-riding” phenomenon in peer to peer file exchange networks. We have demonstrated that our approach allows to drive the system to a state where most peers share files, despite given initial conditions that are supposed to drive the system to a state where no peer shares. We have also executed experiments with different configurations of the architecture to identify the different means to improve the performance of the architecture.

6.2.2. Multi Modeling and multi-simulation

Participants: Vincent Chevrier, Christine Bourjot, Benjamin Camus.

Laurent Ciarletta (Madynes team, LORIA) is an external collaborator.

Complex systems generally require to use different points of view (abstraction levels) at the same time on the system in order to capture and to understand all the dynamics and the complexity. Being made of different interacting parts, a model of a complex system also requires simultaneously modeling and simulation (M&S) tools from different scientific fields.

We proposed the AA4MM meta-model [56] is to build a society of models, simulators and simulation softwares that solves the core challenges of multimodelling and simulation coupling in an homogeneous perspective.

This year we focused on systems that naturally involve entities at different levels of description: micro and macro levels with their dynamics and and their articulations : emergence (upward causation, from micro to macro levels) and immergence (downward causation, from macro to micro levels). We relied on Bourgine's generic view of the relationship between complex phenomenon's levels and their temporal evolution [50]. We proposed an extension of the AA4MM concepts[13] in order to adapt them to emergence and immergence specifications. A simple example of multi-level modeling of a flocking phenomenon has been implemented to illustrate our proposal.

6.2.3. Robustness of Cellular Automata and Reactive Multi-Agent Systems

Participants: Olivier Bouré, Vincent Chevrier, Nazim Fatès.

Our research on emergent collective behaviours focuses on robustness analysis, that is the behavioural resistance to perturbations in collective systems. We progressed in the knowledge of how to tackle this issue in the case of cellular automata (CA) and multi-agent systems (MAS).

The density classification problem was taken as a simple example for studying how decentralised computations can be carried out with simple cells. Although it is known that this problem can not be solved perfectly, we derived analytic calculations to understand how stochastic cellular automata provide good solutions [3]. A collaboration with mathematicians lead us to study how to extend this result to the infinite-space case [25] and to the 2D finite case [19].

Two papers resulting from the Amybia projects were published : experimental results on phase transitions obtained with FPGAs [7] and the description on a robotics experiment that demonstrates the robustness of a bio-inspired aggregation method [5].

The results on asynchronous information transmission in cellular automata were consolidated [2]. Original definitions of asynchronism were also developed in lattice-gas cellular automata [11], which allows us to complete our spectrum of models for which robustness can be studied analytically and with numerical simulations.

6.2.4. Robotics Systems and Ambiant Intelligence

6.2.4.1. Robotics systems : autonomy, cooperation, robustness

6.2.4.1.1. Local control based platooning

Participants: Alexis Scheuer, Olivier Simonin, François Charpillet, Jano Yazbeck.

We consider decentralised control methods to operate autonomous vehicles at close spacings to form a platoon. We study models inspired by the flocking approach, where each vehicle computes its control from its local perceptions. We investigate different decentralised models in order to provide robust and scalable solutions. Open questions concern collision avoidance, stability and multi-platoon navigation.

In order to reduce the tracking error (*i.e.* the distance between each follower's path and the path of its predecessor), we developed both an innovative approach [58] and a new lateral control law. This lateral control law reduces the tracking error faster than other existing control laws. This control law, and the experimental results obtained with it, has been submitted to 2013 IEEE International Conference on Robotics and Automation. Its integration with a previously defined secure longitudinal control law [55] has also been studied, and will be submitted soon to 2013 IFAC Intelligent Autonomous Vehicles Symposium.

6.2.4.1.2. Adaptation of autonomous vehicle traffic to perturbations

Participants: Mohamed Tlig, Olivier Simonin, Olivier Buffet.

In the context of the European project InTraDE, the problem studied in the context of Mohamed Tlig's PhD thesis is to handle the displacements of numerous IAVs (Intelligent Autonomous Vehicles) in a seaport. Here we assume a supervisor planning the routes of the vehicles in the port. However, in such a large and complex system, different unexpected events can arise and degrade the traffic : failure of a vehicle, human mistake while driving, obstacle on roads, local re-planning, and so on.

We started focusing on a first important sub-problem of space resource sharing among multiple agents: how to ensure the crossing of two opposed flows of vehicles on a road when one of the two paths is blocked by an obstacle. To overcome this problem, blocked vehicles have to coordinate with vehicles of the other side to share the road and manage delays. The objective is to improve traffic flow and reduce the emergence of traffic jam. After formalizing this problem, we have defined and studied in simulation two decision rules that produce two different strategies: the first one alternates between two vehicles from each side of the road, and the second one gives priority to the vehicle with the highest delay. This work has been presented in ICTAI'12 [29].

We are now considering more complex situations, e.g., when multiple flows of vehicles share more than one crossroad.

6.2.4.1.3. Multi-robot exploration and mapping : The Carotte Challenge

Participants: Olivier Simonin, François Charpillet, Antoine Bautin, Nicolas Beaufort.

In the context of the ANR/DGA Carotte Challenge, we study since 2009 new strategies and algorithms for multi-robot exploration and mapping. The proposed models are experimented with real autonomous mobile robots at LORIA and every year at the Carotte challenge. Our consortium, called "Cart-o-matic", is composed of members from Université d'Angers (LISA) and from Maia team-project (our industrial partner has left the consortium in 2011).

The year 2012 produced several results :

- In June, **we won the final edition of the Carotte challenge !** This result was obtained in particular by the efficiency and the robustness of the multi-robot strategy we proposed. Our system also provided one of the best map of the contest.
- We developed a software platform, including SLAM, Planning and multi-robot explorations algorithms. These softwares have been protected by copyrights (APP), see 5.4 .
- We presented the results in different publications : RIA revue [8], ICIRA'2012 International Conference [10] (Finalist for the Best student paper).
- Antoine Bautin wrote his PhD thesis, that he will defend in the beginning of year 2013. This work proposes new frontier assignation algorithms for multi-robot exploration. We defined a new heuristics, based on counting the robots towards a frontier rather than considering only the distance between robots and frontiers. For these purpose we developed algorithms based on wavefronts computations (artificial potential fields).We measured on benchmarks that our algorithm outperforms the two classical approaches *closest frontier* and *Greedy assignation*.
- In Oct. 2012, Nassim Kaldé started a PhD thesis (MENRT scholarship), advised by F. Charpillet and O. Simonin. We aim at continuing the work of the Cartomatic project, under new hypothesis and constrains on communications and complexity of the environment to explore.

6.2.4.2. Intelligent environments and health assistance

6.2.4.2.1. Spatial computing: iTiles network

Participants: Olivier Simonin, François Charpillet, Lionel Havet, Mihai Andries.

Olivier Rochel (Inria research engineer, SED Nancy) is an external collaborator.

In the context of ambient intelligence and robotic assistance, we explore the definition of an active floor composed of connected nodes, forming a network of cells. We consider different way of computation, as spatial calculus, to define robust and self-adaptive functions in the environment. We aim at dealing with walk analysis, surveillance of people activity (actimetry) and assistance (control of assistant robots, etc.).

This work can be summarized in several points :

- We asked Hikob company to design the iTile model we defined at the end of year 2011. In 2012, a network of 90 iTiles has been installed on the floor of the smart apartment of the center. This apartment is an experimental platform developed in the context of the “Situated Computer Science” Action of the CPER MISN (Lorraine region, Inria and government fundings). See InfoSitu.
- Each iTile is composed of one node connected to embedded sensors and to its neighboring tiles. A tile holds 4 weight sensors, an accelerometer and 16 LEDs. A simulator of the iTile network has been developed by Olivier Rochel. This tools makes easier the development on the real tiles.
- Several functions have been developed and are currently under experiments: (i) detection of a person walking on the floor (ii) tracking of feet position (iii) propagation and display of information in the network.
- We are involved since 2010 in the PAL Inria large scale initiative (Personally Assisted Living). In this context, Mihai Andries started a PhD thesis in oct. 2012 (funded by Inria-PAL). This PhD. aims at studying the iTiles model and its possibility for assistance functions. We also study models allowing robots to interact and to use the iTile network.

6.2.4.2.2. Center of pressure and Step Detection of a person walking on our intelligent floor

Participants: Amandine Dubois, François Charpillet.

It is quite easy to estimate in realtime the center of pressure of a person walking on the intelligent floor described above. From a sequence of center of pressure, we conceived a system categorizing the set of measures into two sets :

- foot: the measure belongs to the pressure trace left by a foot on the floor,
- transition: the center of pressure corresponds to what happens when the person passes his right leg or left from backwards to forwards.

This has been done in a first time using an heuristic algorithm and then using an HMM. From this categorization it's then easy to estimate classical gait parameters such as length of the steps or speed of the walk.

6.2.4.2.3. Pose estimation of several kinects

Participants: Nicolas Beaufort, François Charpillet.

Tracking one or several persons using several Kinects required to solved the calibration, i.e estimation of the pose of each kinect in the scene, knowing that the area covered by each Depth camera don't overlap with other (because of interference). We have addressed this issue using a SLAM approach implemented within a GPU.

6.2.4.2.4. Fall prevention and Fall detection

Participants: Amandine Dubois, François Charpillet.

A major problem of public health is the loss of autonomy of elderly people usually caused by the falls. Since 2003 one of the goal of MAIA team is to develop a system allowing to detect falls and also to analyze the gait deterioration to prevent falls. A first approach consisted in developing a markerless human motion capture system estimating the 3D positions of the body joints over time. This system used a dynamic Bayesian network and a factored particle filtering algorithm. Since 2011, we used a new approach using Microsoft Kinect camera which allows to acquire at the same time a RGB and a depth image to deal of the problem of the gait. After the extraction of the human from the background, we calculate the gait parameters from the center of mass of a person. Some parameters, as the length and the time of steps, the speed of the gait, allow to predict a deterioration of the gait of a person and an increase of the risk of falls [17].

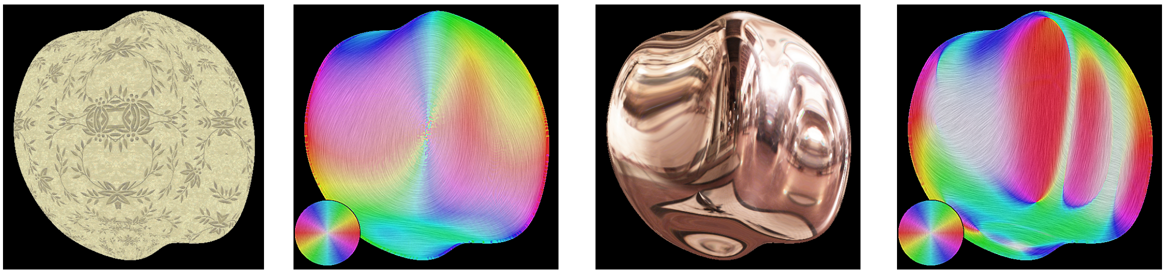
Another use of the extraction of center of mass of a person from the Kinect camera is to determine the activity of a person. The method uses a Hidden Markov Model to distinguish eight activities of the daily life (sitting, walking, lying (on a couch, on a bed), lying down, falling, going up on the obstacles, squatting and bending). We set up an experiment in a smart room to validate our results. Concerning the gait parameters we compare them to the real values obtained making the young subjects wake with pads soaked with ink under the shoes on the paper. The results show that there is a difference of 3-4cm between length provided by our Kinect algorithm and the real length provided by the paper. Concerning the detection of the activity, we ask to 28 subjects to perform eight situations (corresponding to the eight states of the HMM). The results showed that each situation is recognized exept “bending”, falls are detected correctly and there are no false positives except “sitting” and “qqsquatting” which are detected instead of “bending”.

MANAO Team

5. New Results

5.1. Axis 1:Analysis and Simulation

5.1.1. First Order Analysis of Shading



Texuring | 1st order gradient field | Environment reflection | 2st order gradient field

Figure 8. First-oder analysis [21] have shown that shading variations are caused by depth variations (first-order gradient field) and by normal variations (second-order fields). These fields are visualized using hue and saturation to indicate direction and magnitude of the flow respectively.

We introduced [21] a novel method for producing convincing pictures of shaded objects based entirely on 2D image operations. This approach, which we call image-based shading design, offers direct artistic control in the picture plane by deforming image primitives so that they appear to conform to specific 3D shapes. Using a differential analysis of reflected radiance, we have identified the two types of surface flows involved in the depiction of shaded objects, which are consistent with recent perceptual studies. We have also introduced two novel deformation operators that closely mimic surface flows while providing direct artistic controls in real-time.

5.1.2. Rational BRDF

Over the last two decades, much effort has been devoted to accurately measuring Bidirectional Reflectance Distribution Functions (BRDFs) of real-world materials and to use efficiently the resulting data for rendering. Because of their large size, it is difficult to use directly measured BRDFs for real-time applications, and fitting the most sophisticated analytical BRDF models is still a complex task. In this paper, we introduce Rational BRDF [19], a general-purpose and efficient representation for arbitrary BRDFs, based on Rational Functions (RFs). Using an adapted parametrization, we demonstrate how Rational BRDFs offer 1) a more compact and efficient representation using low-degree RFs, 2) an accurate fitting of measured materials with guaranteed control of the residual error, and 3) efficient importance sampling by applying the same fitting process to determine the inverse of the Cumulative Distribution Function (CDF) generated from the BRDF for use in Monte-Carlo rendering.

5.2. Axis 2: From Acquisition to Display

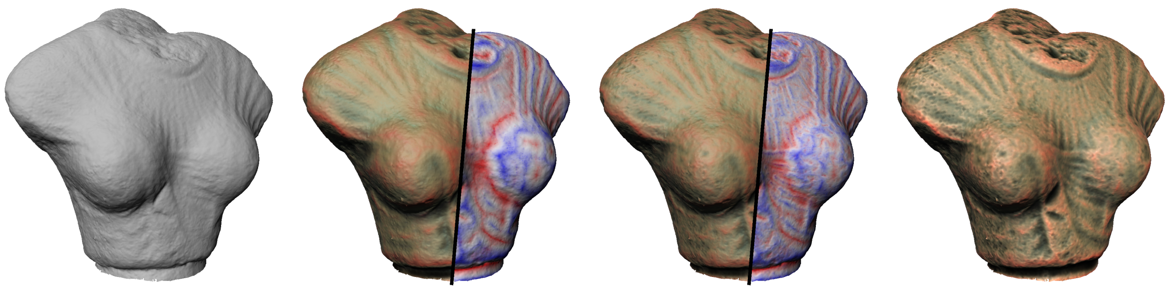
5.2.1. Outdoor Lighting for Augmented Reality

Figure 9. Consistent illumination of a virtual car in real outdoor lighting.

In augmented reality, one of the key tasks to achieve a convincing visual appearance consistency between virtual objects and video scenes is to have a coherent illumination along the whole sequence. As outdoor illumination is largely dependent on the weather, the lighting condition may change from frame to frame. We have proposed [17] a full image-based approach for online tracking of outdoor illumination variations from videos captured with moving cameras. Our key idea is to estimate the relative intensities of sunlight and skylight via a sparse set of planar feature-points extracted from each frame. To address the inevitable feature misalignments, a set of constraints are introduced to select the most reliable ones. Exploiting the spatial and temporal coherence of illumination, the relative intensities of sunlight and skylight are finally estimated by using an optimization process. We have validated our technique on a set of real-life videos and show that the results with our estimations are visually coherent along the video sequences (cf. Figure 9).

5.3. Axis 3: Rendering, Visualization and Illustration

5.3.1. Surface Relief Analysis for Illustrative Shading

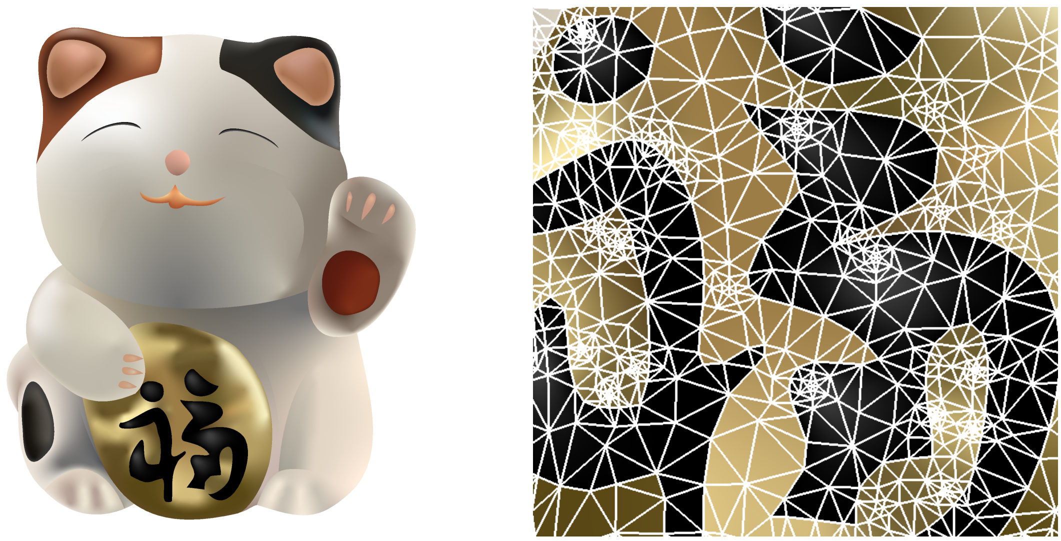


(a) (b) (c) (d)

Figure 10. Given a detailed surface (a), we analyze its relief to locate relief features in the neighborhood of each surface point (b). We focus on three types of features: convexities, concavities, and inflexions, shown on the right half with blue, red and white colors respectivelly. Extracted information is used to assign them different shading functions: here we use three different lit-spheres, shown on the left half. An additional accessibility shading effect helps convey relief cavities. Features are extracted and combined at multiple scales to depict relevant relief details (c). Finally, radiance scaling is added to enhance the relief based on the curvature at each feature (d).

Rendering techniques are often used to convey shape in scientific illustrations. We present an analysis technique that leverages the complexity found in detailed 3D models for illustrative shading purposes. Given a smooth base surface with relief, it locates relief features (concavities, convexities and inflections) around each surface point and at multiple scales, using cubic-polynomial fitting. This object-space, per-vertex information is then used to guide a variety of shading techniques including normal enhancement, feature visualization, accessibility shading and radiance scaling. Thanks to this approach, features at multiple scales are easily combined, filtered and shaded, allowing users to explore surface relief in real-time (cf. Figure 10).

5.4. Axis 4: Editing and Modeling



(a) (b)

Figure 11. A complex image obtained using our vectorial solver (a), with a close-up view showing the automatically generated intermediate triangle mesh (b).

5.4.1. Free form vector gradients

The creation of free-form vector drawings as been greatly improved in recent years with techniques based on harmonic or bi-harmonic interpolation. Such methods offer the best trade-off between sparsity (keeping the number of control points small) and expressivity (achieving complex shapes and gradients). Unfortunately, the lack of a robust and versatile method to compute such images still limits their use in real-world applications. We developed a vectorial solver for the computation of free-form vector gradients based on a non-conform Finite Element Methods (FEM). Its key feature is to output a low-level vector representation suitable for very fast GPU accelerated rasterization and close-form evaluation (fig. 11). This intermediate representation is hidden from the user: it is dynamically updated using FEM during drawing when control points are edited. We demonstrated novel usages of vector drawings such as instancing, layering, deformation, texture and environment mapping. Finally, we also generalized and extended the set of drawing possibilities, in particular, by showing how to locally control vector gradients. This work has been published at SIGGRAPH Asia [16] and featured by the 3DFV website [24].

5.4.2. Growing Least Squares (GLS) for the Analysis of Manifolds in Scale-Space

We created a novel approach for the multi-scale analysis of point-sampled manifolds of co-dimension 1. It is based on a variant of Moving Least Squares, whereby the evolution of a geometric descriptor at increasing scales is used to locate pertinent locations in scale-space, hence the name “Growing Least Squares (GLS)”. Compared to existing scale-space analysis methods, our approach is the first to provide a continuous solution in space and scale dimensions, without requiring any parametrization, connectivity or uniform sampling. An important implication is that we identify multiple pertinent scales for any point on a manifold, a property that had not yet been demonstrated in the literature. In practice, our approach exhibits an improved robustness to change of input, and is easily implemented in a parallel fashion on the GPU, and it can be used in a wide variety of applications. For example, the GLS can be used for the detection of similarity, according to a given scale range (see Figure 12). This work has been published at the Symposium of Geometry Processing [18].

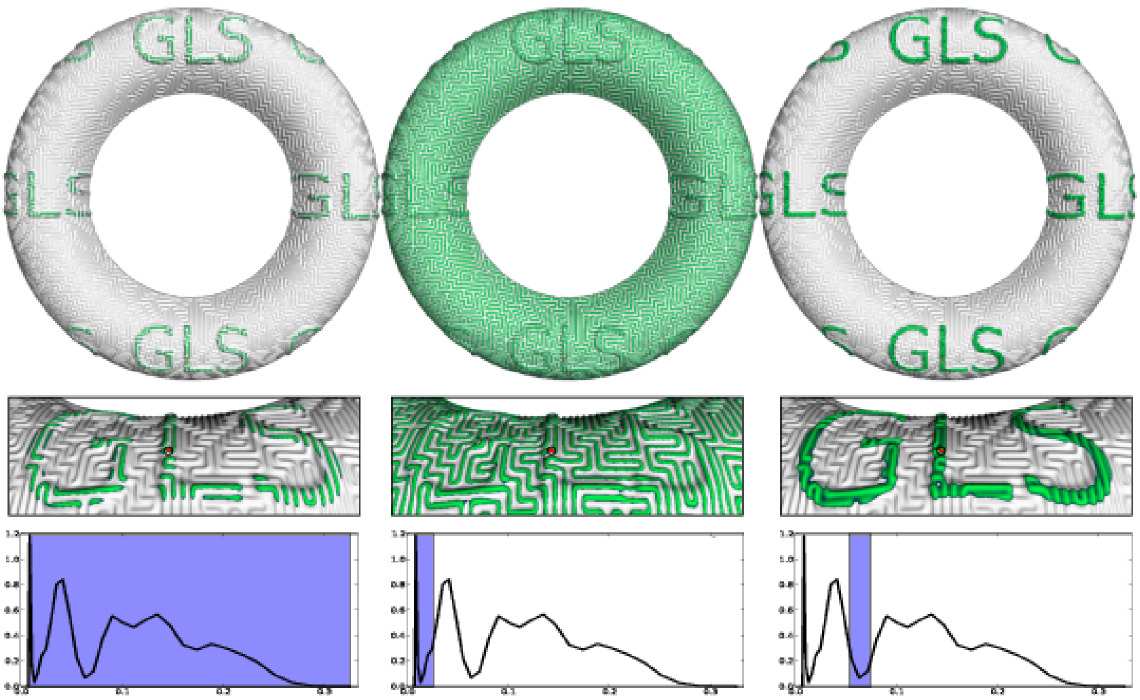


Figure 12. GLS Multi-scale similarity. Top and middle rows: For a selected point (in red), similar points are selected (in green) via our dissimilarity measure. The similarity is computed for each vertex and interpolated per fragment during the rendering. Bottom row: the type of selected feature depends on a user-controlled global prior (shown as a blue box), which is locally refined by our geometric variation. In (a), all scales are selected. In (b), only the fine displacement pattern emerges. In (c), the large-scale GLS letters are properly segmented.

MAVERICK Team

6. New Results

6.1. Computer visualization

6.1.1. Immersive Virtual Environment for Visuo-Vestibular Therapy: Preliminary Results

Participants: Jean-Dominique Gascuel, Henri Payno, Sébastien Schmerber, Olivier Martin.

The sense of equilibrium aggregates several interacting cues. On patients with vestibular loss, vision plays a major role. In this study, the goal is to propose a new immersive therapy based on 3D opto-kinetic stimulation. We propose to demonstrate that 3D monoscopic optical flows are an efficient tool to stimulate adaptive postural adjustment. We developed an immersive therapeutic platform that enables to tune the balance task difficulty by managing optic flow speed and gaze anchoring (Figure 5). METHODOLOGY: the immersive sessions proposed to vestibular areflexic patients are composed of a repetition of dynamic optic flows, with varying speed and presence or not of a gaze anchor. The balance adjustments are recorded by a force plate, and quantified by the length of the center of pressure trajectory. RESULTS: Preliminary analysis shows that (i) Patients report a strong immersion felling in the motion flow, triggering more intense motor response to "fight against fall" than in standard opto-kinetic protocols; (ii) An ANOVA factorial design shows a significant effect of flow speed, session number and gaze anchor impact. CONCLUSION: This study shows that 3D immersive stimulation removes essential limits of traditional opto-kinetic stimulators (limited 2D motions and remaining fixed background cues). Moreover, the immersive optic flow stimulation is an efficient tool to induce balance adaptive reactions in vestibular patients. Hence, such a platform appears to be a powerful therapeutic tool for training and relearning of balance control processes.

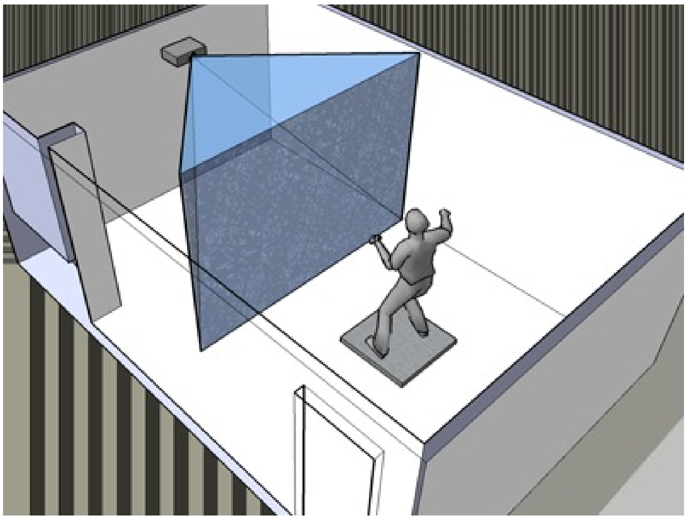


Figure 5. The immersive platform, installed in an available room of the hospital. The large retro projected screen is at 60 cm of the patient, covering most of its visual field. The patient is standing on a force plate, recording CoP.

6.1.2. Evaluation of Depth of Field for Depth Perception in DVR

Participants: Pascal Grosset, Mathias Schott, Georges-Pierre Bonneau, Hansen Charles.

we present a user study on the use of Depth of Field for depth perception in Direct Volume Rendering (Figure 6). Direct Volume Rendering with Phong shading and perspective projection is used as the baseline. Depth of Field is then added to see its impact on the correct perception of ordinal depth. Accuracy and response time are used as the metrics to evaluate the usefulness of Depth of Field. The on site user study has two parts: static and dynamic. Eye tracking is used to monitor the gaze of the subjects. From our results we see that though Depth of Field does not act as a proper depth cue in all conditions, it can be used to reinforce the perception of which feature is in front of the other. The best results (high accuracy & fast response time) for correct perception of ordinal depth is when the front feature (out of the users were to choose from) is in focus and perspective projection is used.

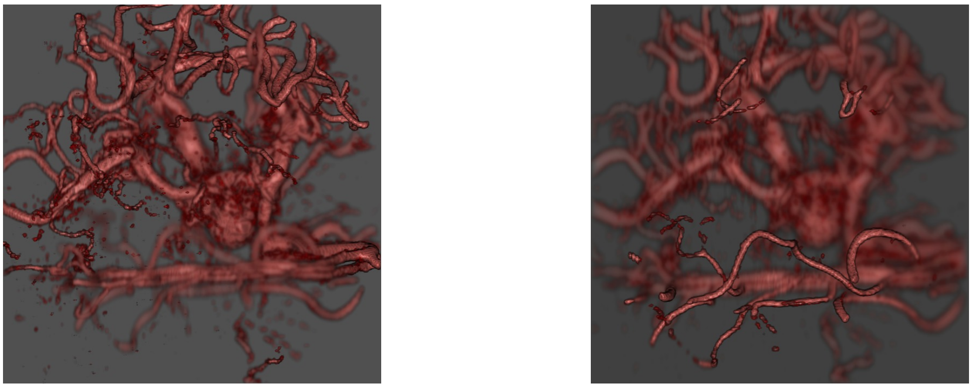


Figure 6. Aneurism. Depth of Field reinforces the perception of which feature is in front of the other.

6.1.3. Volume Preserving FFD for Programmable Graphics Hardware

Participants: Stefanie Hahmann, Georges-Pierre Bonneau, Sébastien Barbier, Gershon Elber, Hans Hagen.

Free Form Deformation (FFD) is a well established technique for deforming arbitrary object shapes in space. Although more recent deformation techniques have been introduced, amongst them skeleton-based deformation and cage based deformation, the simple and versatile nature of FFD is a strong advantage, and justifies its presence in nowadays leading commercial geometric modeling and animation software systems. Since its introduction in the late 80's, many improvements have been proposed to the FFD paradigm, including control lattices of arbitrary topology, direct shape manipulation and GPU implementation. Several authors have addressed the problem of volume preserving FFD. These previous approaches either make use of expensive non-linear optimization techniques, or resort to first order approximation suitable only for small-scale deformations. In this paper we take advantage from the multi-linear nature of the volume constraint in order to derive a simple, exact and explicit solution to the problem of volume preserving FFD. Two variants of the algorithm are given, without and with direct shape manipulation. Moreover, the linearity of our solution enables to implement it efficiently on GPU (Figure 7).

6.1.4. Sharp feature preserving MLS surface reconstruction based on local feature line approximations

Participants: Christopher Weber, Stefanie Hahmann, Hans Hagen, Georges-Pierre Bonneau.

Sharp features in manufactured and designed objects require particular attention when reconstructing surfaces from unorganized scan point sets using moving least squares (MLS) fitting. It's an inherent property of MLS fitting that sharp features are smoothed out. Instead of searching for appropriate new fitting functions our approach computes a modified local point neighborhood so that a standard MLS fitting can be applied

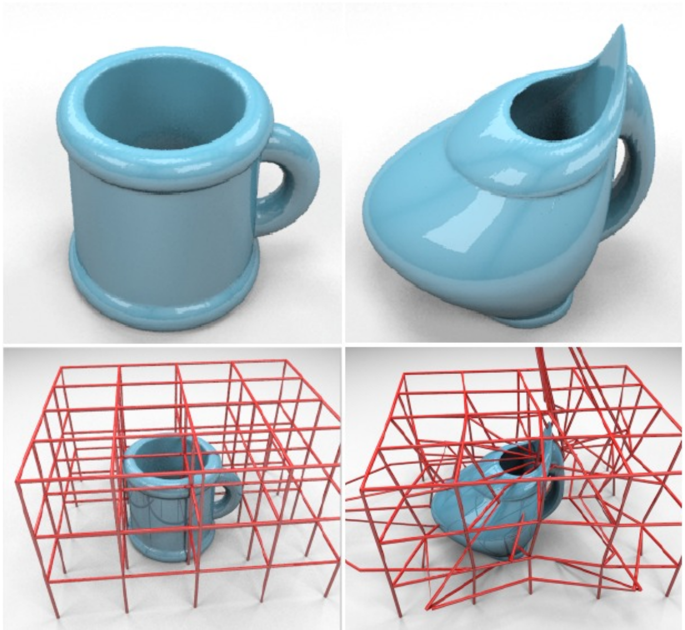


Figure 7. Sculptured cup with volume preservation.

enhanced by sharp features reconstruction. We present a two-stage algorithm. In a pre-processing step sharp feature points are marked first. This algorithm is robust to noise since it is based on Gauss map clustering. In the main phase, the selected feature points are used to locally approximate the feature curve and to segment and enhance the local point neighborhood. The MLS projection thus leads to a piecewise smooth surface preserving all sharp features. The method is simple to implement and able to preserve line-type features as well as corner-type features during reconstruction (Figure 8).

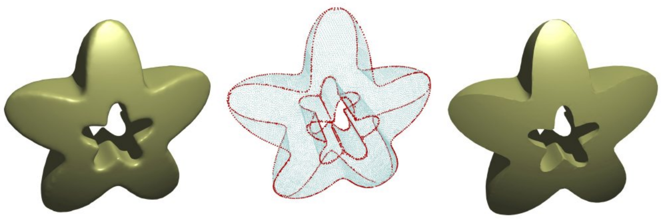


Figure 8. Left: standard MLS surface. Middle: feature point detection in point cloud. Right: sharp feature preserving MLS.

6.2. Expressive rendering

6.2.1. Active Strokes: Coherent Line Stylization for Animated 3D Models

Participants: Pierre Bénard, Lu Jingwan, Forrester Cole, Adam Finkelstein, Joëlle Thollot.

We present a method for creating coherently animated line drawings that include strong abstraction and stylization effects (Figure 9). These effects are achieved with active strokes: 2D contours that approximate and track the lines of an animated 3D scene. Active strokes perform two functions: they connect and smooth unorganized line samples, and they carry coherent parameterization to support stylized rendering. Line samples are approximated and tracked using active contours ("snakes") that automatically update their arrangment and topology to match the animation. Parameterization is maintained by brush paths that follow the snakes but are independent, permitting substantial shape abstraction without compromising fidelity in tracking. This approach renders complex models in a wide range of styles at interactive rates, making it suitable for applications like games and interactive illustrations.

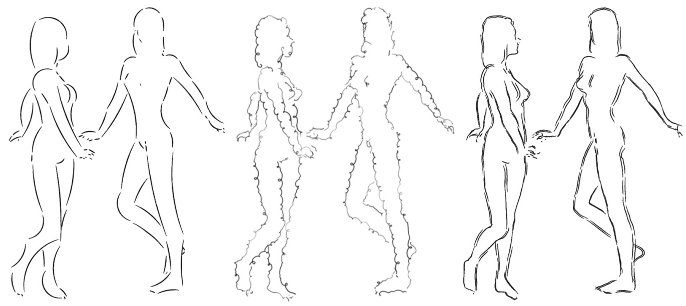


Figure 9. Stylization examples. Woman in two poses and three styles: arcs, loopy offsets, and overdrawn.

6.2.2. Temporally Coherent Video Stylization

Participants: Pierre Bénard, Joëlle Thollot, John Collomosse.

The transformation of video clips into stylized animations remains an active research topic in Computer Graphics. A key challenge is to reproduce the look of traditional artistic styles whilst minimizing distracting flickering and sliding artifacts; i.e. with temporal coherence. This chapter surveys the spectrum of available video stylization techniques, focusing on algorithms encouraging the temporally coherent placement of rendering marks, and discusses the trade-offs necessary to achieve coherence. We begin with flow-based adaptations of stroke based rendering (SBR) and texture advection capable of painting video. We then chart the development of the field, and its fusion with Computer Vision, to deliver coherent mid-level scene representations. These representations enable the rotoscoping of rendering marks on to temporally coherent video regions, enhancing the diversity and temporal coherence of stylization. In discussing coherence, we formalize the problem of temporal coherence in terms of three defined criteria, and compare and contrast video stylization using these.

6.3. Illumination simulation

6.3.1. Accurate fitting of measured reflectances using a Shifted Gamma micro-facet distribution

Participants: Mahdi M. Bagher, Cyril Soler, Nicolas Holzschuch.

Material models are essential to the production of photo-realistic images. Measured BRDFs provide accurate representation with complex visual appearance, but have larger storage cost. Analytical BRDFs such as Cook- Torrance provide a compact representation but fail to represent the effects we observe with measured appearance. Accurately fitting an analytical BRDF to measured data remains a challenging problem. In this paper we introduce the SGD micro-facet distribution for Cook-Torrance BRDF. This distribution accurately models the behavior of most materials. As a consequence, we accurately represent all measured BRDFs using a single lobe. Our fitting procedure is stable and robust, and does not require manual tweaking of the parameters (Figure 10).

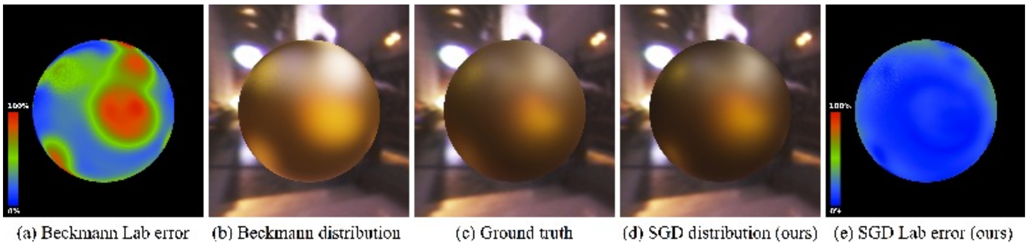


Figure 10. fitting of measured reflectances: comparison between ground truth and our approach.

6.3.2. Interactive rendering of acquired materials on dynamic geometry using bandwidth prediction

Participants: Mahdi M. Bagher, Cyril Soler, Kartic Subr, Laurent Belcour, Nicolas Holzschuch.

Shading complex materials such as acquired reflectances in multi-light environments is computationally expensive. Estimating the shading integral involves sampling the incident illumination independently at several pixels. The number of samples required for this integration varies across the image, depending on an intricate combination of several factors. Adaptively distributing computational budget across the pixels for shading is therefore a challenging problem. In this paper we depict complex materials such as acquired reflectances, interactively, without any precomputation based on geometry. We first estimate the approximate spatial and angular variation in the local light field arriving at each pixel. This *local bandwidth* accounts for combinations of a variety of factors: the reflectance of the object projecting to the pixel, the nature of the illumination, the local geometry and the camera position relative to the geometry and lighting. We then exploit this bandwidth information to adaptively sample for reconstruction and integration. For example, fewer pixels per area are shaded for pixels projecting onto diffuse objects, and fewer samples are used for integrating illumination incident on specular objects (Figure 11).

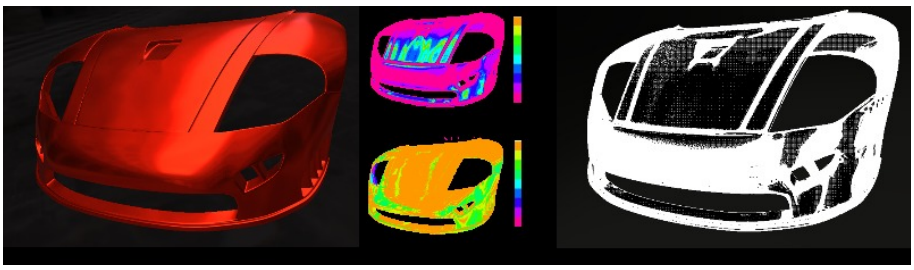


Figure 11. Interactive rendering of aquired materials. Center: predicted bandwidth and variance. Right: sample points where we compute illumination. Left: rendered result.

6.3.3. Real-Time Rendering of Rough Refraction

Participants: Charles De Rousiers, Adrien Bousseau, Kartic Subr, Nicolas Holzschuch, Ravi Ramamoorthi.

We present an algorithm to render objects made of transparent materials with rough surfaces in real-time, under all-frequency distant illumination (Figure 12). Rough surfaces cause wide scattering as light enters and exits objects, which significantly complicates the rendering of such materials. We present two contributions to approximate the successive scattering events at interfaces, due to rough refraction: First, an approximation of the Bidirectional Transmittance Distribution Function (BTDF), using spherical Gaussians, suitable for real-time estimation of environment lighting using pre-convolution; second, a combination of cone tracing and macro-geometry filtering to efficiently integrate the scattered rays at the exiting interface of the object. We demonstrate the quality of our approximation by comparison against stochastic ray-tracing. Furthermore we propose two extensions to our method for supporting spatially varying roughness on object surfaces and local lighting for thin objects.

(a) Ground truth (b) Our method

Figure 12. Real-Time Rendering of Rough Refraction

6.3.4. Multiple-scattering and double-scattering effects in translucent materials

Participants: Jean-Dominique Gascuel, Nicolas Holzschuch.

Some materials, such as coffee, milk or marble, have a soft translucent aspect because of sub-surface scattering: light enters them, is scattered several times inside before leaving in a different place. A full representation of sub-surface scattering effects in illumination simulation is computationally expensive. The main difficulty comes from multiple scattering events: the high number of events increases the incertainty on the result, forcing us to allocate more time for the computations. In this paper, we show that there is a strong correlation between the surface effects of multiple scattering inside the material and the effects after just two scatter events. This knowledge will help for accelerating multiple scattering effects.We also provide a model for fast computation of double-scattering events, using a precomputed density function we store in a compact way (Figure 13).

6.3.5. Frequency analysis of participating media

Participants: Laurent Belcour, Cyril Soler, Kavita Bala.

Computing global illumination in participating media is frustratingly expensive: while the computation itself is long and complicated, the result involve very smooth regions of illumination. This motivates an a priori approach to find out how fast the resulting image will vary in space (i.e. it’s spatial frequency) to adapt computation effort to reach the maximal efficiency. For this we are extending the theory of Fourier Analysis of

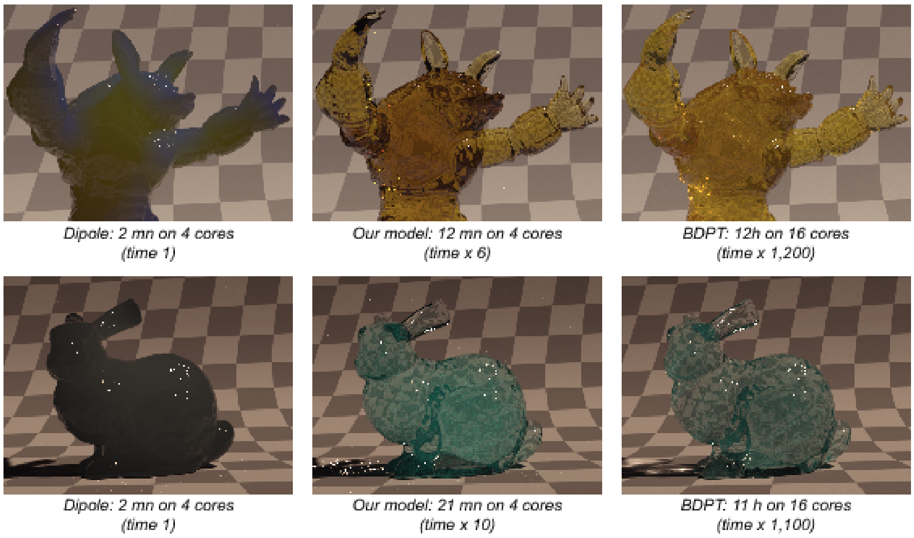


Figure 13. Multiple-scattering and double-scattering effects in translucent materials

light transport to participating media. Our work builds on the covariance analysis of light transport developed by Laurent Belcour in his PhD Thesis. It offers the possibility to drastically accelerate the algorithms involved in the computation of the illumination in scenes with participating media (Figure 14).

6.4. Complex scenes

6.4.1. A Survey of Non-linear Pre-filtering Methods for Efficient and Accurate Surface Shading

Participants: Eric Bruneton, Fabrice Neyret.

Rendering a complex surface accurately and without aliasing requires the evaluation of an integral for each pixel, namely a weighted average of the outgoing radiance over the pixel footprint on the surface. The outgoing radiance is itself given by a local illumination equation as a function of the incident radiance and of the surface properties. Computing all this numerically during rendering can be extremely costly. For efficiency, especially for real-time rendering, it is necessary to use precomputations. When the fine scale surface geometry, reflectance and illumination properties are specified with maps on a coarse mesh (such as color maps, normal maps, horizon maps or shadow maps), a frequently used simple idea is to pre-filter each map linearly and separately. The averaged outgoing radiance, i.e., the average of the values given by the local illumination equation is then estimated by applying this equation to the averaged surface parameters. But this is really not accurate because this equation is non-linear, due to self-occlusions, self-shadowing, non-linear reflectance functions, etc. Some methods use more complex pre-filtering algorithms to cope with these non-linear effects. This paper is a survey of these methods. We start with a general presentation of the problem of pre-filtering complex surfaces. We then present and classify the existing methods according to the approximations they make to tackle this difficult problem. Finally, an analysis of these methods allows us to highlight some generic tools to pre-filter maps used in non-linear functions, and to identify open issues to address the general problem.

6.4.2. Real-time Realistic Rendering and Lighting of Forests

Participants: Eric Bruneton, Fabrice Neyret.

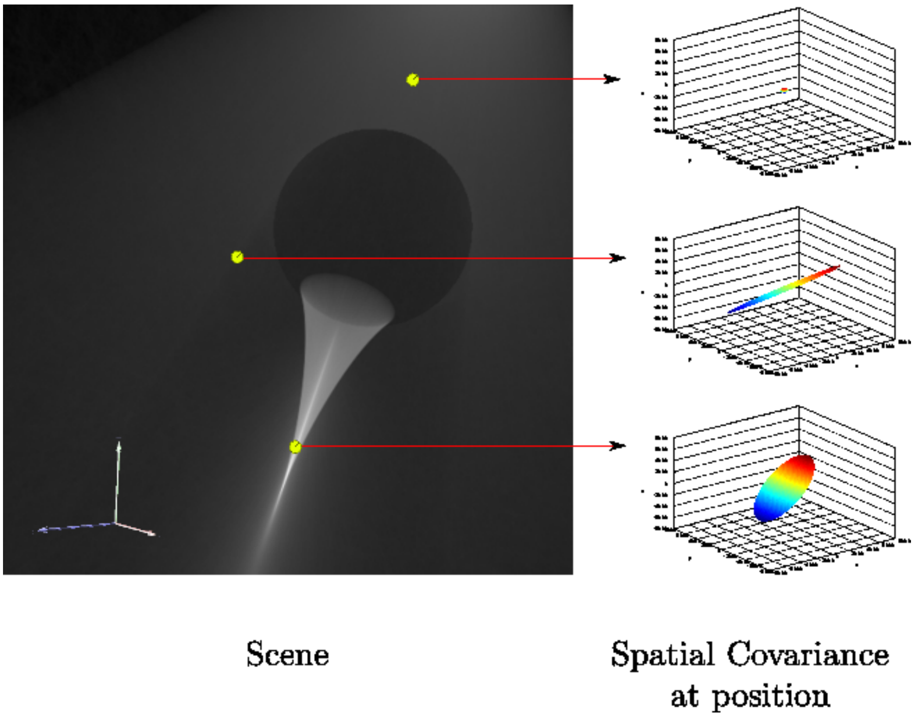


Figure 14. A glass sphere castinc a volumetric caustic in participating media with multiple scattering. We can predict the covariance of the spectrum of the illumination locally everywhere in the volume so as to adapt the computation effort.

Realistic real-time rendering and lighting of forests is an important aspect for simulators and video games. This is a difficult problem, due to the massive amount of geometry: aerial forest views display millions of trees on a wide range of distances, from the camera to the horizon. Light interactions, whose effects are visible at all scales, are also a problem: sun and sky dome contributions, shadows between trees, inside trees, on the ground, and view-light masking correlations. In this paper we present a method to render very large forest scenes in realtime, with realistic lighting at all scales, and without popping nor aliasing (Figure 15). Our method is based on two new forest representations, z-fields and shader-maps, with a seamless transition between them. Our first model builds on light fields and height fields to represent and render the nearest trees individually, accounting for all lighting effects. Our second model is a location, view and light dependent shader mapped on the terrain, accounting for the cumulated subpixel effects. Qualitative comparisons with photos show that our method produces realistic results.

Figure 15. Some real-time results obtained with our method, showing large forest scenes with a wide range of view distances, various tree densities and lighting conditions.

6.4.3. Representing Appearance and Pre-filtering Subpixel Data in Sparse Voxel Octrees

Participants: Eric Heitz, Fabrice Neyret.

Sparse Voxel Octrees (SVOs) represent efficiently complex geometry on current GPUs. Despite the fact that LoDs come naturally with octrees, interpolating and filtering SVOs are still issues in current approaches. In this paper, we propose a representation for the appearance of a detailed surface with associated attributes stored within a voxel octree. We store macro- and micro-descriptors of the surface shape and associated attributes in each voxel. We represent the surface macroscopically with a signed distance field and we encode subvoxel microdetails with Gaussian descriptors of the surface and attributes within the voxel. Our voxels form a continuous field interpolated through space and scales, through which we cast conic rays. Within the ray marching steps, we compute the occlusion distribution produced by the macro-surface inside a pixel footprint, we use the microdescriptors to reconstruct light- and view-dependent shading, and we combine fragments in an A-buffer way. Our representation efficiently accounts for various subpixel effects. It can be continuously interpolated and filtered, it is scalable, and it allows for efficient depth-of-field. We illustrate the quality of these various effects by displaying surfaces at different scales, and we show that the timings per pixel are scale-independent (Figure 16).

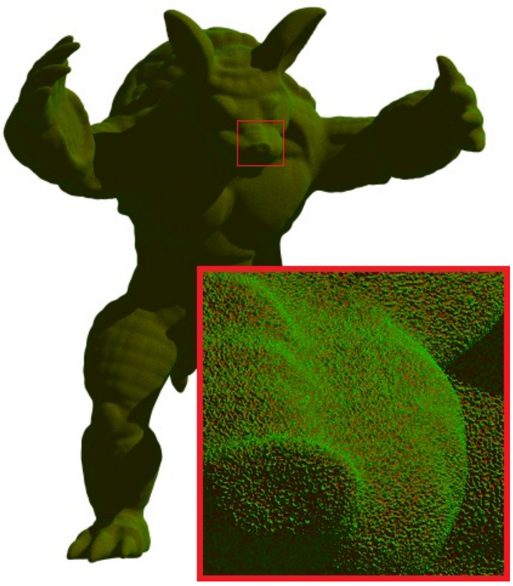


Figure 16. Our method allows for correct filtering of color variations, like anti-aliasing demonstrated here.

METISS Project-Team

6. New Results

6.1. Audio and speech content processing

Audio segmentation, speech recognition, motif discovery, audio mining

6.1.1. Audio motif discovery

Participants: Frédéric Bimbot, Laurence Catanese.

This work was performed in close collaboration with Guillaume Gravier from the Texmex project-team.

As an alternative to supervised approaches for multimedia content analysis, where predefined concepts are searched for in the data, we investigate content discovery approaches where knowledge emerge from the data. Following this general philosophy, we pursued work on motif discovery in audio contents.

Audio motif discovery is the task of finding out, without any prior knowledge, all pieces of signals that repeat, eventually allowing variability. The developed algorithms allows discovering and collecting occurrences of repeating patterns in the absence of prior acoustic and linguistic knowledge, or training material.

Former work extended the principles of seeded discovery to near duplicate detection and spoken document retrieval from examples [41].

In 2012, the work achieved consisted in consolidating previously obtained results with the motif discovery algorithm and making implementation choices regardless of the structure and the code, in order to minimize the computation time. This has lead to the creation of a software prototype called MODIS.

After the code has been thoroughly optimised, further optimizations to improve the system performances was to change the method used for the search of similarities between patterns. A new functionality has been added to get rid of unrelevant patterns like silence in speech. New versions of dynamic time warping have been implemented, as well as the possibility to downsample the input sequence during the process, which allows a huge gain of computation time.

The Inria/Metiss team has participated to the IRIT P5 evaluation for repetitive musical motifs discovery. The motif discovery software has been adapted to respect the input and output format defined for the task. The run has been made on a evaluation corpus comprised of French radio broadcast from YACAST.

This work has been carried out in the context of the Quaero Project.

6.1.2. Landmark-driven speech recognition

Participant: Stefan Ziegler.

This work is supervised by Guillaume Gravier and Bogdan Ludusan from the Texmex project-team.

Our previous studies indicate that acoustic-phonetic approaches to ASR, while they cannot achieve state-of-the-art ASR performance by themselves, can prevent HMM-based ASR from degrading, by integrating additional knowledge into the decoding.

In our previous framework we inserted knowledge into the decoding by detecting time frames (referred to as landmarks) which estimate the presence of the active broad phonetic class. This enables the use of a modified version of the viterbi decoding that favours states that are coherent with the detected phonetic knowledge[65].

In 2012 we focused on two major issues. First, we aimed at finding new ways to model and detect phonetic landmarks. Our second focus was on the extension of our landmark detector towards a full acoustic-phonetic framework, to model speech by a variety of articulatory features.

Our new approach for the classification and detection of speech units focuses on developping landmark-models that are different from existing frame-based approaches to landmark detection[64]. In our approach, we use segmentation to model any time-variable speech unit by a fixed-dimensional observation vector. After training any desired classifier, we can estimate the presence of a desired speech unit by searching for each time frame the corresponding segment, that provides the maximum classification score.

We used this segment-based landmark-detection inside a standalone acoustic-phonetic framework that models speech as a stream of articulatory features. In this framework we first search for relevant broad phonetic landmarks, before attaching each landmark with the full set of articulatory features.

Integrating these articulatory feature streams into a standard HMM-based speech recognizer by weighted linear combination improves speech recognition up to 1.5

Additionally, we explored the possibilities of using stressed syllables as an information to guide the viterbi decoding. This work was carried under the leaderhip of Bogdan Ludusan from the team TEXMEX at IRISA [56].

6.1.3. Speech-driven functionalities for interactive television

Participants: Grégoire Bachman, Guylaine Le Jan, Nathan Souviraà-Labastie, Frédéric Bimbot.

In the context of the collaborative ReV-TV project, the Metiss research group has contributed to technological solutions for the demonstration of new concepts of interactive television, integrating a variety of modalities (audio/voice, gesture, image, haptic feed-back).

The focus has been to provide algorithmic solutions to some advanced audio processing and speech recognition tasks, in particular : keywords recognition, lip synchronisation for an avatar, voice emotion recognition and interactive vocal control.

The main challenges adressed in the project have been to robustify state-of-the-art based technologies to the diversity of adverse conditions, to provide real-time response and to ensure the smooth integration of the various interactive technologies involved in the project.

The work of the project has resulted in a demonstration which was presented at the Forum Imagina 2012

6.2. Recent results on sparse representations

Sparse approximation, high dimension, scalable algorithms, dictionary design, graph wavelets

The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the field of sparse representations, which is at the core of the FET-Open European project (FP7) SMALL (Sparse Models, Algorithms and Learning for Large-Scale Data, see section 8.2.1.1), the ANR project ECHANGE (ECHantillonnage Acoustique Nouvelle GEnération, see section 8.1.1.2), and the ERC project PLEASE (projections, Learning and Sparsity for Efficient Data Processing, see section 8.2.1.2).

6.2.1. A new framework for sparse representations: analysis sparse models

Participants: Rémi Gribonval, Sangnam Nam, Nancy Bertin, Srdjan Kitic.

Main collaboration: Mike Davies, Mehrdad Yaghoobi (Univ. Edinburgh), Michael Elad (The Technion).

In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of *few* columns from a given matrix (the dictionary). An alternative *analysis-based* model can be envisioned, where an analysis operator multiplies the signal, leading to a *cosparse* outcome. Within the SMALL project, we initiated a research programme dedicated to this analysis model, in the context of a generic missing data problem (e.g., compressed sensing, inpainting, source separation, etc.). We obtained a uniqueness result for the solution of this problem, based on properties of the analysis operator and the measurement matrix. We also considered a number of pursuit algorithms for solving the missing data problem, including an L1-based and a new greedy method called GAP (Greedy Analysis Pursuit). Our simulations demonstrated the appeal of the analysis model, and the success of the pursuit techniques presented.

These results have been published in conferences and in a journal paper [42]. Other algorithms based on iterative cosparse projections [83] as well as extensions of GAP to deal with noise and structure in the cosparse representation have been developed, with applications to toy MRI reconstruction problems and acoustic source localization and reconstruction from few measurements [58].

6.2.2. Theoretical results on sparse representations and dictionary learning

Participants: Rémi Gribonval, Sangnam Nam, Nancy Bertin.

Main collaboration: Karin Schnass (EPFL), Mike Davies (University of Edinburgh), Volkan Cevher (EPFL), Simon Foucart (Université Paris 5, Laboratoire Jacques-Louis Lions), Charles Soussen (Centre de recherche en automatique de Nancy (CRAN)), Jérôme Idier (Institut de Recherche en Communications et en Cybernétique de Nantes (IRCCyN)), Cédric Herzet (Equipe-projet FLUMINANCE (Inria - CEMAGREF, Rennes)), Morten Nielsen (Department of Mathematical Sciences [Aalborg]), Gilles Puy, Pierre Vandergheynst, Yves Wiaux (EPFL), Mehrdad Yaghoobi, Rodolphe Jenatton, Francis Bach (Equipe-projet SIERRA (Inria, Paris)), Boaz Ophir, Michael Elad (Technion), Mark D. Plumbley (Queen Mary, University of London).

Sparse recovery conditions for Orthogonal Least Squares : We pursued our investigation of conditions on an overcomplete dictionary which guarantee that certain ideal sparse decompositions can be recovered by some specific optimization principles / algorithms. We extended Tropp's analysis of Orthogonal Matching Pursuit (OMP) using the Exact Recovery Condition (ERC) to a first exact recovery analysis of Orthogonal Least Squares (OLS). We showed that when ERC is met, OLS is guaranteed to exactly recover the unknown support. Moreover, we provided a closer look at the analysis of both OMP and OLS when ERC is not fulfilled. We showed that there exist dictionaries for which some subsets are never recovered with OMP. This phenomenon, which also appears with ℓ_1 minimization, does not occur for OLS. Finally, numerical experiments based on our theoretical analysis showed that none of the considered algorithms is uniformly better than the other. This work has been submitted for publication in a journal [86]. More recently, we obtained simpler coherence-based conditions [85].

Performance guarantees for compressed sensing with spread spectrum techniques : We advocate a compressed sensing strategy that consists of multiplying the signal of interest by a wide bandwidth modulation before projection onto randomly selected vectors of an orthonormal basis. Firstly, in a digital setting with random modulation, considering a whole class of sensing bases including the Fourier basis, we prove that the technique is universal in the sense that the required number of measurements for accurate recovery is optimal and independent of the sparsity basis. This universality stems from a drastic decrease of coherence between the sparsity and the sensing bases, which for a Fourier sensing basis relates to a spread of the original signal spectrum by the modulation (hence the name "spread spectrum"). The approach is also efficient as sensing matrices with fast matrix multiplication algorithms can be used, in particular in the case of Fourier measurements. Secondly, these results are confirmed by a numerical analysis of the phase transition of the l1-minimization problem. Finally, we show that the spread spectrum technique remains effective in an analog setting with chirp modulation for application to realistic Fourier imaging. We illustrate these findings in the context of radio interferometry and magnetic resonance imaging. This work has been accepted for publication in a journal [45].

Dictionary learning : An important practical problem in sparse modeling is to choose the adequate dictionary to model a class of signals or images of interest. While diverse heuristic techniques have been proposed in the litterature to learn a dictionary from a collection of training samples, there are little existing results which provide an adequate mathematical understanding of the behaviour of these techniques and their ability to recover an ideal dictionary from which the training samples may have been generated.

In 2008, we initiated a pioneering work on this topic, concentrating in particular on the fundamental theoretical question of the identifiability of the learned dictionary. Within the framework of the Ph.D. of Karin Schnass, we developed an analytic approach which was published at the conference ISCCSP 2008 [13] and allowed us to describe "geometric" conditions which guarantee that a (non overcomplete) dictionary is "locally identifiable" by ℓ^1 minimization.

In a second step, we focused on estimating the number of sparse training samples which is typically sufficient to guarantee the identifiability (by ℓ^1 minimization), and obtained the following result, which is somewhat surprising considering that previous studies seemed to require a combinatorial number of training samples to guarantee the identifiability: the local identifiability condition is typically satisfied as soon as the number of training samples is roughly proportional to the ambient signal dimension. The outline of the second result was published in conferences [12], [25]. These results have been published in the journal paper [15].

Analysis Operator Learning for Overcomplete Cosparse Representations : Besides standard dictionary learning, we also considered learning in the context of the cosparse model. We consider the problem of learning a low-dimensional signal model from a collection of training samples. The mainstream approach would be to learn an overcomplete dictionary to provide good approximations of the training samples using sparse synthesis coefficients. This famous sparse model has a less well known counterpart, in analysis form, called the cosparse analysis model. In this new model, signals are characterized by their parsimony in a transformed domain using an overcomplete analysis operator. We consider two approaches to learn an analysis operator from a training corpus.

The first one uses a constrained optimization program based on L1 optimization. We derive a practical learning algorithm, based on projected subgradients, and demonstrate its ability to robustly recover a ground truth analysis operator, provided the training set is of sufficient size. A local optimality condition is derived, providing preliminary theoretical support for the well-posedness of the learning problem under appropriate conditions. Extensions to deal with noisy training samples are currently investigated, and a journal paper is under revision [87].

In the second approach, analysis "atoms" are learned sequentially by identifying directions that are orthogonal to a subset of the training data. We demonstrate the effectiveness of the algorithm in three experiments, treating synthetic data and real images, showing a successful and meaningful recovery of the analysis operator.

Connections between sparse approximation and Bayesian estimation: Penalized least squares regression is often used for signal denoising and inverse problems, and is commonly interpreted in a Bayesian framework as a Maximum A Posteriori (MAP) estimator, the penalty function being the negative logarithm of the prior. For example, the widely used quadratic program (with an ℓ^1 penalty) associated to the LASSO / Basis Pursuit Denoising is very often considered as MAP estimation under a Laplacian prior in the context of additive white Gaussian noise (AWGN) reduction.

A first result, which we published last year, highlights the fact that, while this is *one* possible Bayesian interpretation, there can be other equally acceptable Bayesian interpretations. Therefore, solving a penalized least squares regression problem with penalty $\phi(x)$ need not be interpreted as assuming a prior $C \cdot \exp(-\phi(x))$ and using the MAP estimator. In particular, we showed that for *any* prior P_X, the minimum mean square error (MMSE) estimator is the solution of a penalized least square problem with some penalty $\phi(x)$, which can be interpreted as the MAP estimator with the prior $C \cdot \exp(-\phi(x))$. Vice-versa, for *certain* penalties $\phi(x)$, the solution of the penalized least squares problem is indeed the MMSE estimator, with a certain prior P_X. In general $dP_X(x) \neq C \cdot \exp(-\phi(x))dx$.

A second result, obtained in collaboration with Prof. Mike Davies and Prof. Volkan Cevher (a paper is under revision) characterizes the "compressibility" of various probability distributions with applications to underdetermined linear regression (ULR) problems and sparse modeling. We identified simple characteristics of probability distributions whose independent and identically distributed (iid) realizations are (resp. are not) compressible, i.e., that can be approximated as sparse. We prove that many priors which MAP Bayesian interpretation is sparsity inducing (such as the Laplacian distribution or Generalized Gaussian distributions with exponent p<=1), are in a way inconsistent and do not generate compressible realizations. To show this, we identify non-trivial undersampling regions in ULR settings where the simple least squares solution outperform oracle sparse estimation in data error with high probability when the data is generated from a sparsity inducing prior, such as the Laplacian distribution [39].

6.3. Emerging activities on compressive sensing, learning and inverse problems

Compressive sensing, acoustic wavefields, audio inpainting,

6.3.1. Nearfield acoustic holography (ECHANGE ANR project)

Participants: Rémi Gribonval, Nancy Bertin.

Main collaborations: Albert Cohen (Laboratoire Jacques-Louis Lions, Université Paris 6), Laurent Daudet, Gilles Chardon, François Ollivier, Antoine Peillot (Institut Jean Le Rond d'Alembert, Université Paris 6)

Compressed sensing is a rapidly emerging field which proposes a new approach to sample data far below the Nyquist rate when the sampled data admits a sparse approximation in some appropriate dictionary. The approach is supported by many theoretical results on the identification of sparse representations in overcomplete dictionaries, but many challenges remain open to determine its range of effective applicability. METISS has chosen to focus more specifically on the exploration of Compressed Sensing of Acoustic Wavefields, and we have set up the ANR collaborative project ECHANGE (ECHantillonnage Acoustique Nouvelle GEnération) which began in January 2009. Rémi Gribonval is the coordinator of the project.

In 2010, the activity on ECHANGE has concentrated on Nearfield acoustic holography (NAH), a technique aiming at reconstructing the operational deflection shapes of a vibrating structure, from the near sound field it generates. In this application scenario, the objective is either to improve the quality of the reconstruction (for a given number of sensors), or reduce the number of sensors, or both, by exploiting a sparsity hypothesis which helps regularizing the inverse problem involved.

Contributions of the team in this task spans: notations and model definitions, experimental setting design and implementation, choice of an adapted dictionary in which the sparsity hypothesis holds, improved acquisition strategies through pseudo-random sensor arrays and/or spatial multiplexing of the inputs, experimental study of robustness issues, and theoretical study of potential success guarantees based on the restricted isometry property (which revealed being not verified in our case, despite improved experimental performance).

A paper about robustness issues and spatial multiplexing (an alternative to building antennas with random sensor position) was published in GRETSI last year and as a journal paper this year [38].

6.3.2. Sparse reconstruction for underwater acoustics (ECHANGE ANR project)

Participants: Rémi Gribonval, Nancy Bertin.

Main collaborations: Jacques Marchal, Pierre Cervenka (UPMC Univ Paris 06)

Underwater acoustic imaging is traditionally performed with beamforming: beams are formed at emission to insonify limited angular regions; beams are (synthetically) formed at reception to form the image. We proposed to exploit a natural sparsity prior to perform 3D underwater imaging using a newly built flexible-configuration sonar device. The computational challenges raised by the high-dimensionality of the problem were highlighted, and we described a strategy to overcome them. As a proof of concept, the proposed approach was used on real data acquired with the new sonar to obtain an image of an underwater target. We discussed the merits of the obtained image in comparison with standard beamforming, as well as the main challenges lying ahead, and the bottlenecks that will need to be solved before sparse methods can be fully exploited in the context of underwater compressed 3D sonar imaging. This work has been published in [61] and a journal paper is in preparation.

6.3.3. Audio inpainting (SMALL FET-Open project)

Participants: Rémi Gribonval, Nancy Bertin, Corentin Guichaoua.

Main collaborations: Amir Adler, Michael Elad (Computer Science Department, The Technion, Israel); Maria G. Jafari, Mark D. Plumbley (Centre for Digital Music, Department of Electronic Engineering, Queen Mary University of London, U.K.).

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. It consists in reconstructing a set of missing pixels in an image based on the observation of the remaining pixels. Sparse representations have proved to be particularly appropriate to address this problem. However, inpainting audio data has never been defined as such so far.

METISS has initiated a series of works about audio inpainting, from its definition to methods to address it. This research has begun in the framework of the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL (Sparse Models, Algorithms and Learning for Large-Scale data) which began in January 2009. Rémi Gribonval is the coordinator of the project. The research on audio inpainting has been conducted by Valentin Emiya in 2010 and 2011.

The contributions consist of:

- defining audio inpainting as a general scheme where missing audio data must be estimated: it covers a number of existing audio processing tasks that have been addressed separately so far – click removal, declipping, packet loss concealment, unmasking in time-frequency;
- proposing algorithms based on sparse representations for audio inpainting (based on Matching Pursuit and on l_1 minimization);
- addressing the case of audio declipping (*i.e.* desaturation): thanks to the flexibility of our inpainting algorithms, they can be constrained so as to include the structure of signals due to clipping in the objective to optimize. The resulting performance are significantly improved. This work will appear as a journal paper [33].

Current and future works deal with developping advanced sparse decomposition for audio inpainting, including several forms of structured sparsity (*e.g.* temporal and multichannel joint-sparsity), dictionary learning for inpainting, and several applicative scenarios (declipping, time-frequency inpainting).

6.3.4. Blind Calibration of Compressive Sensing systems

Participants: Rémi Gribonval, Cagdas Bilen.

Main collaborations: Gilles Chardon, Laurent Daudet (Institut Langevin), Gilles Puy (EPFL)

We consider the problem of calibrating a compressed sensing measurement system under the assumption that the decalibration consists in unknown gains on each measure. We focus on blind calibration, using measures performed on a few unknown (but sparse) signals. A naive formulation of this blind calibration problem, using l1 minimization, is reminiscent of blind source separation and dictionary learning, which are known to be highly non-convex and riddled with local minima. In the considered context, we show that in fact this formulation can be exactly expressed as a convex optimization problem, and can be solved using off-the-shelf algorithms. Numerical simulations demonstrate the effectiveness of the approach even for highly uncalibrated measures, when a sufficient number of (unknown, but sparse) calibrating signals is provided. We observe that the success/failure of the approach seems to obey sharp phase transitions. This work has been published at ICASSP 2012 [54], and an extension dealing with the problem of phase-only decalibration, using techniques revolving around low-rank matrix recovery, has been submitted to ICASSP 2013. A journal version is in preparation.

6.3.5. Compressive Gaussian Mixture estimation

Participants: Rémi Gribonval, Anthony Bourrier.

Main collaborations: Gilles Blanchard (University of Potsdam), Patrick Perez (Technicolor R&D, FR)

When fitting a probability model to voluminous data, memory and computational time can become prohibitive. In this paper, we pro- pose a framework aimed at fitting a mixture of isotropic Gaussians to data vectors by computing a low-dimensional sketch of the data. The sketch represents empirical moments of the underlying probability distribution. Deriving a reconstruction algorithm by analogy with compressive sensing, we experimentally show that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough. Our algorithm provides good reconstruction and scales to higher dimensions than previous probability mixture estimation algorithms, while consuming less memory in the case of numerous data. It also provides a privacy-preserving data analysis tool, since the sketch does not disclose information about individual datum it is based on. This work has been submitted for publication at ICASSP 2013.

6.3.6. Nearest neighbor search for arbitrary kernels with explicit embeddings

Participants: Rémi Gribonval, Anthony Bourrier.

Main collaborations: Hervé Jégou (TEX-MEX team), Patrick Perez (Technicolor R&D, FR)

Many algorithms have been proposed to handle efficient search in large databases for simple metrics such as the Euclidean distance. However, few approaches apply to more sophisticated Positive Semi-Definite (PSD) kernels. In this document, we propose for such kernels to use the concept of explicit embedding and to cast the search problem into a Euclidean space. We first describe an exact nearest neighbor search technique which relies on bounds on the approximation of the kernel. We show that, in the case of SIFT descriptors, one can retrieve the nearest neighbor with probability 1 by computing only a fraction of the costly kernels between the query and the database vectors. We then propose to combine explicit embedding with a recent Euclidean approximate nearest neighbor search method and show that it leads to significant improvements with respect to the state-of-the-art methods which rely on an implicit embedding. The database vectors being indexed by short codes, the approach is shown to scale to a dataset comprising 200 million vectors on a commodity server. This work has been submitted for journal publication [74]

6.4. Music Content Processing and Music Information Retrieval

Acoustic modeling, non-negative matrix factorisation, music language modeling, music structure

6.4.1. Music language modeling

Participants: Frédéric Bimbot, Dimitris Moreau, Stanisław Raczyński, Emmanuel Vincent.

Main collaboration: S. Fukayama (University of Tokyo, JP)

Music involves several levels of information, from the acoustic signal up to cognitive quantities such as composer style or key, through mid-level quantities such as a musical score or a sequence of chords. The dependencies between mid-level and lower- or higher-level information can be represented through acoustic models and language models, respectively.

We pursued our pioneering work on music language modeling, with a particular focus on the joint modeling of "horizontal" (sequential) and "vertical" (simultaneous) dependencies between notes by log-linear interpolation of the corresponding conditional distributions. We identified the normalization of the resulting distribution as a crucial problem for the performance of the model and proposed an exact solution to this problem [81]. We also applied the log-linear interpolation paradigm to the joint modeling of melody, key and chords, which evolve according to different timelines [80]. In order to synchronize these feature sequences, we explored the use of beat-long templates consisting of several notes as opposed to short time frames containing a fragment of a single note.

The limited availability of multi-feature symbolic music data is currently an issue which prevents the training of the developed models on sufficient amounts of data for the unsupervised probabilistic approach to significantly outperform more conventional approaches based on musicological expertise. We outlined a procedure for the semi-automated collection of large-scale multifeature music corpora by exploiting the wealth of music data available on the web (audio, MIDI, leadsheets, lyrics, etc) together with algorithms for the automatic detection and alignment of matching data. Following this work, we started collecting pointers to data and developing such algorithms.

6.4.2. Music structuring

Participants: Frédéric Bimbot, Gabriel Sargent, Emmanuel Vincent.

External collaboration: Emmanuel Deruty (as an independant consultant)

The structure of a music piece is a concept which is often referred to in various areas of music sciences and technologies, but for which there is no commonly agreed definition. This raises a methodological issue in MIR, when designing and evaluating automatic structure inference algorithms. It also strongly limits the possibility to produce consistent large-scale annotation datasets in a cooperative manner.

This year, our methodology for the *semiotic* annotation of music pieces has developed [72] and concretized into a set of principles, concepts and conventions for locating the boundaries and determining metaphoric labels of music segments [53] [71]. The method relies on a new concept for characterizing the inner organization of music segments called the System & Contrast (S&C) model [73]. At the time of writing this text, the annotation of over 400 music pieces is being finalized and will be released to the MIR scientific community.

In parallel to this work aiming at specifying the task of music structure description, we have designed, implemented and tested new algorithms for segmenting and labeling music into structural units. The segmentation process is formulated as a cost optimization procedure, accounting for two terms : the first one corresponds to the characterization of structural segments by means of the fusion of audio criteria, whereas the second term relies on a regularity constraint on the resulting segmentation. Structural labels are estimated as a probabilistic automaton selection process. A recent development of this work has included the S&C model in the algorithm.

Different systems based on these principles have been tested in the context of the Quaero Project and the MIREX international evaluation campaigns in 2010, 2011 and 2012 (see for instance [66], in 2012).

6.5. Source separation

Source separation, sparse representations, probabilistic model, source localization

6.5.1. A general framework for audio source separation

Participants: Frédéric Bimbot, Rémi Gribonval, Nobutaka Ito, Emmanuel Vincent.

Main collaborations: H. Tachibana (University of Tokyo, JP), N. Ono (National Institute of Informatics, JP)

Source separation is the task of retrieving the source signals underlying a multichannel mixture signal. The state-of-the-art approach consists of representing the signals in the time-frequency domain and estimating the source coefficients by sparse decomposition in that basis. This approach relies on spatial cues, which are often not sufficient to discriminate the sources unambiguously. Recently, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [44], which generalizes a number of existing techniques including our former study on spectral GMMs [34]. This framework makes it possible to quickly design a new model adapted to the data at hand and estimate its parameters via the EM algorithm. As such, it is expected to become the basis for a number of works in the field, including our own.

Since the EM algorithm is sensitive to initialization, we devoted a major part of our work to reducing this sensitivity. One approach is to use some prior knowledge about the source spatial covariance matrices, either via probabilistic priors [75] or via deterministic subspace constraints [76]. The latter approach was the topic of the PhD thesis of Nobutaka Ito who defended this year [30]. A complementary approach is to initialize the parameters in a suitable way using source localization techniques specifically designed for environments involving multiple sources and possibly background noise [37].

6.5.2. Exploiting filter sparsity for source localization and/or separation

Participants: Alexis Benichoux, Emmanuel Vincent, Rémi Gribonval, Frédéric Bimbot.

Main collaboration: Simon Arberet (EPFL)

Estimating the filters associated to room impulse responses between a source and a microphone is a recurrent problem with applications such as source separation, localization and remixing.

We considered the estimation of multiple room impulse responses from the simultaneous recording of several known sources. Existing techniques were restricted to the case where the number of sources is at most equal to the number of sensors. We relaxed this assumption in the case where the sources are known. To this aim, we proposed statistical models of the filters associated with convex log-likelihoods, and we proposed a convex optimization algorithm to solve the inverse problem with the resulting penalties. We provided a comparison between penalties via a set of experiments which shows that our method allows to speed up the recording process with a controlled quality tradeoff. A journal paper including extensive experiments with real data is in preparation.

We also investigated the filter estimation problem in a blind setting, where the source signals are unknown. We proposed an approach for the estimation of sparse filters from a convolutive mixture of sources, exploiting the time-domain sparsity of the mixing filters and the sparsity of the sources in the time-frequency (TF) domain. The proposed approach is based on a wideband formulation of the cross-relation (CR) in the TF domain and on a framework including two steps: (a) a clustering step, to determine the TF points where the CR is valid; (b) a filter estimation step, to recover the set of filters associated with each source. We proposed for the first time a method to blindly perform the clustering step (a) and we showed that the proposed approach based on the wideband CR outperforms the narrowband approach and the GCC-PHAT approach by between 5 dB and 20 dB. This work has been submitted for publication as a journal paper.

On a more theoretical side, we studied the frequency permutation ambiguity traditionnally incurred by blind convolutive source separation methods. We focussed on the filter permutation problem in the absence of scaling, investigating the possible use of the temporal sparsity of the filters as a property enabling permutation correction. The obtained theoretical and experimental results highlight the potential as well as the limits of sparsity as an hypothesis to obtain a well-posed permutation problem. This work has been published in a conference [52] and is accepted for publication as a journal paper, to appear in 2013.

6.5.3. Towards real-world separation and remixing applications

Participants: Nancy Bertin, Frédéric Bimbot, Jules Espiau de Lamaestre, Jérémy Paret, Laurent Simon, Nathan Souviraà-Labastie, Joachim Thiemann, Emmanuel Vincent.

Shoko Araki, Jonathan Le Roux (NTT Communication Science Laboratories, JP)

We participated in the organization of the 2011 Signal Separation Evaluation Campaign (SiSEC) [51], [59]. Following our founding role in the organization of this campaign, we wrote an invited paper summarizing the outcomes of the three first editions of this campaign from 2007 to 2010 [47]. While some challenges remain, this paper highlighted that progress has been made and that audio source separation is closer than ever to successful industrial applications. This is also exemplified by the ongoing i3DMusic project and the recently signed contracts with Canon Research Centre France and MAIA Studio.

In order to exploit our know-how for these real-world applications, we investigated issues such as how to implement our algorithms in real time [60], how to reduce artifacts [40] and how best to exploit extra information or human input. In addition, while the state-of-the-art quality metrics previously developed by METISS remain widely used in the community, we proposed some improvements to the perceptually motivated metrics introduced last year [62].

6.5.4. Source separation for multisource content indexing

Participants: Kamil Adiloğlu, Emmanuel Vincent.

Main collaborations: Jon Barker (University of Sheffield, UK), Mathieu Lagrange (IRCAM, FR), Alexey Ozerov (Technicolor R&D, FR)

Another promising real-world application of source separation concerns information retrieval from multi-source data. Source separation may then be used as a pre-processing stage, such that the characteristics of each source can be separately estimated. The main difficulty is not to amplify errors from the source separation stage through subsequent feature extraction and classification stages. To this aim, we proposed a principled Bayesian approach to the estimation of the uncertainty about the separated source signals [50], [69], [68] and propagated this uncertainty to the features. We then exploited it in the training of the classifier itself, thereby greatly increasing classification accuracy [43].

This work was applied both to singer identification in polyphonic music [55] and to speech and speaker recognition in real-world nonstationary noise environments. In order to motivate further work by the community, we created a new international evaluation campaign on that topic (CHiME) in 2011 and analyzed the outcomes of the first edition [36].

Some work was also devoted to the modeling of similarity between sound events [32].

MIMETIC Team

6. New Results

6.1. Motion Sensing and analysis

Participants: Franck Multon [contact], Richard Kulpa, Anthony Sorel, Edouard Auvinet.

Sensing human activity is a very active field of research, with a wide range of applications ranging from entertainment and serious games to personal ambient living assistance. MimeTIC aims at proposing original methods to process raw motion capture data in order to compute relevant information according to the application.

In personal ambient living monitoring, we have collaborated with University of Montreal, Department of Computer Science and Operations Research (DIRO) which main activity is biomedical engineering. A co-supervised student is addressing two complementary problems: detecting people falling in everyday environment and providing easy-to-use clinical gait analysis systems for early detection of potential risks of falling. In the last decade, gait analysis has become one of the most active research topics in biomedical research engineering partly due to recent developpement of sensors and signal processing devices and more recently depth cameras. The latters can provide real-time distance measurements of moving objects. In this context, we present a new way to reconstruct body volume in motion using multiple active cameras from the depth maps they provide. A first contribution of this paper is a new and simple external camera calibration method based on several plane intersections observed with a low-cost depth camera which is experimentally validated. A second contribution consists in a body volume reconstruction method based on visual hull that is adapted and enhanced with the use of depth information. Preliminary results based on simulations are presented and compared with classical visual hull reconstruction. These results show that as little as three low-cost depth cameras can recover a more accurate 3D body shape than twenty regular cameras (see figure 4).

In entertainment and serious games, the problem is different as we need to accurately now the action performed by the user in order to react in a convenient manner. Collaboration with Artefacto Company enabled us to develop such motion recognition methods in serious games scenarios. Given motion capture data provided by an optical motion capture system lead to large state vectors in which the relevant information is hidden. Mixture of Gaussians is generally used as an input of Hidden Markov Models to recognize a motion according to this raw data. To simplify, features are generally introduced in order to capture the relevant geometrical property of the motion with either general information (such as joint angles or Cartesian positions) or application-specific information. The former type of information has the advantage to be generic but leads to recognizers that are very sensitive to style and morphology variations. Previously, we have proposed a new generic feature based on morphology-independent representation that enables to tackle this problem [28]. We now have explored the robustness of this type of features for early recognition, when using mixture of Gaussians instead of Hidden Markov Models. We have shown that a motion can be recognized when only 50% of the motion is performed. The recognition rate is especially high with this type of feature compared to classical Euler angles and Cartesian data, especially when a new user is performing the motion [6].

6.2. VR and Sports

Participants: Richard Kulpa [contact], Benoit Bideau, Sébastien Brault, Anne-Marie Burns.

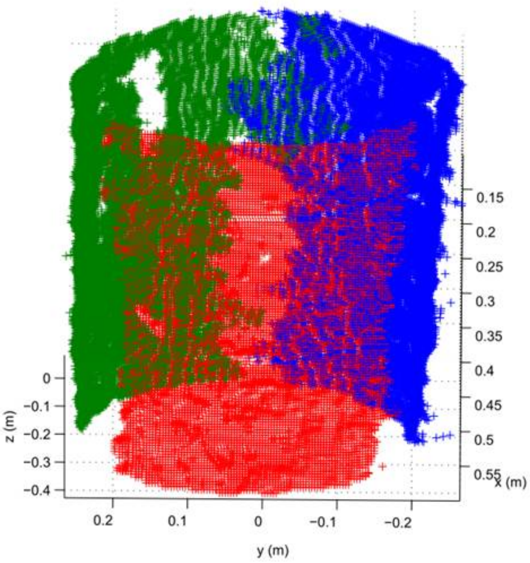


Figure 4. 3D silouhettes reconstructed with three depth-cameras - reconstructed points of a reference cylinder. Each color corresponds to one of the depth camera.

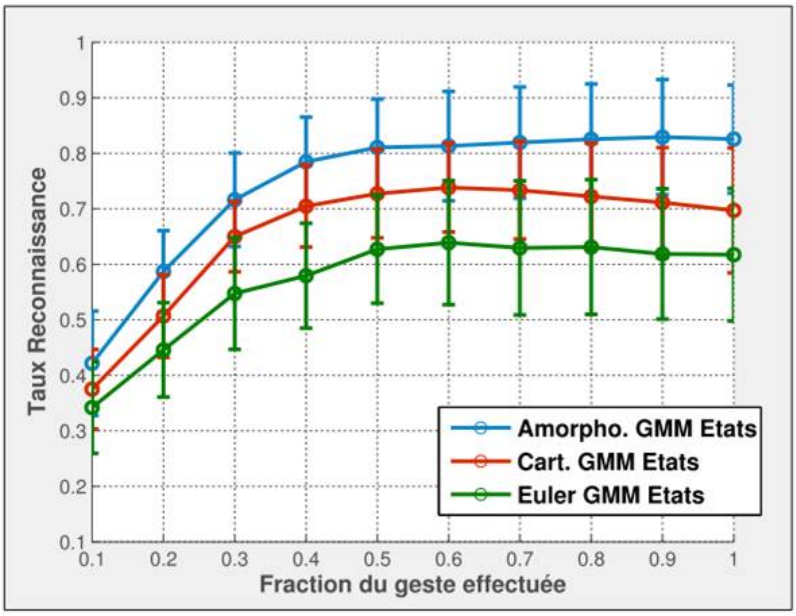


Figure 5. Early recognition of a motion performed by a new user with three different features: Cartesian, Euler and the proposed amorphological features.

In the past, we have worked on the interaction between two opponents in virtual environment. These duels were between a handball goalkeeper and a thrower; and between a rugby defender and an attacker performing deceptive movements. Even if these sports applications are different in terms of kinematic parameters, information picked-up and type of interaction, we have designed a unique framework to simulate such duels in a reality center and to analyze the gestures of real athletes immersed in this environment. This VR framework was validated by showing that behaviors in real and virtual environments were similar. These works have been extended by using perception-action coupling and perception-only studies to evaluate the anticipation of opponents. In order to evaluate the importance of perceived parameters, the ball and/or the character animation was successively hidden to determine their importance and the same kind of study was done on the graphical level of details.

This year, we have addressed the problem of the tennis serve. The first step is the PhD of Caroline Martin who will end next year. This work provides biomechanical analysis of the serve and the influence of the kinematical and dynamic parameters on performance. Thanks to an accepted project funded by the INSEP institute, we are importing this biomechanical model to virtual environment to make perceptual analysis. This work is based on the same methodology used for the detection of deceptive movements in rugby. The next step is to combine the use of cutoffs with biomechanical analysis to extract important kinematic information that could explain differences between experts and novices. This information is then correlated to kinematical parameters of this player. Concurrently, we are working on the creation of models of rugby defenders based on the results of the previous perceptual analyses

Finally, we have worked on the use of virtual environments to train athletes. The first step was to evaluate if a better score in the virtual environment implied only an improvement of the athlete in the virtual game or also a better performance back on the field. The PhD of Anne-Marie Burns has demonstrated that the improvement of training based on virtual environment was similar to training with a real teacher or based on videos. The use of VR for sports training, at least by imitation, is thus possible. Furthermore, we have explored the influence of the self-representation of the immersed learner by displaying his avatar as if he was in front of a virtual mirror. We made both kinematical and evocation analyses. The results do not show significant difference with or without the use of the mirror and it is confirmed by the subjective analysis that shows that the use of the virtual mirror by immersed athletes was limited. This work was partially funded by the Biofeedback project.

6.3. Biomechanics and Motion Analysis

6.3.1. Interaction strategies between two walkers to avoid collision

Participants: Armel Crétual, Julien Pettré, Anne-Hélène Olivier, Antoine Marin.

Walkers are extremely efficient in avoiding collisions, even in relatively condition of density. We experimentally addressed two questions. What are the conditions for walkers to perform adaptations to their trajectory, and second, how avoidance performed in time. We checked several hypothesis, that led to two contributions, as presented in [15]. First, human are able to anticipate the future conditions of an interactions and the distance they would meet. They react accordingly, i.e., if and only if a future risk of collision can be predicted. Second, we demonstrated that the avoidance is performed with anticipation, i.e., avoidance maneuvers are over before walkers get at closest distance.

6.3.2. Quantification of pathological motion

Participant: Armel Crétual.

In clinical routine, precise quantification of patients' gesture remains a challenge. Several simple means are daily used by practicians in physical medecine. Their main drawback is often a large inter-operator variability and even sometimes an intra-operator one. To overcome this, we have developed and validated still simple to remain usable) but much more objective tools in two different fields: gait and shoulder laxity.

First, we have proposed a new index of gait quantification based on EMG profiles called KeR-EGI (for Kerpape-Rennes EMG-based Gait Index). Our recent works allowed us to demonstrate its reproducibility even in patients with severe troubles. Moreover, we have also demonstrated the complementarity of this index based on muscular activation and an index based on kinematics, the Edinburgh Visual Gait Score (EVGS) that can be computed easily from a simple video recordings of the patient's gait. Indeed, we have shown that the relationship between these indices depends on the fact that pathology is congenital or acquired. Using both indices at the time, allows to evaluate the potential kinematics compensation the patient does to improve his/her gait despite a damaged motor control.

Secondly, in shoulder surgery, the surgeon has to choose between different protocols depending on whether the patient is hyperlax or not. Until now, shoulder laxity is very roughly evaluated without actual measurement and above all mobilizing only one axis (external rotation) of this complex joint. By measuring precisely the whole Range Of Motion of 28 subjects recruited to ensure a large spectrum of laxity (from hypo to hyper-laxity), we have shown that the usual clinical indices fail to actually classify subjects, as they do focus on only one dimension of mobility. From, that result, we have then proposed a new method to evaluate laxity that remains simple and usable in daily routine but that takes into account all dimensions of shoulder's mobility.

6.3.3. Modeling gesture in sports: fin swimming

Participants: Nicolas Bideau, Guillaume Nicolas, Benoit Bideau, Richard Kulpa.

In swimming, experimental approaches are commonly used to analyze performance. However, due to obvious limitations in experimental approaches (impoossibility to standardize any situations etc.), it is difficult to characterize surrounding fluid. To overcome this limitation, we currently develop analysis, modeling and simulation of aquatic locomotion, using CFD computer simulation and new methods based on animation of virtual characters.

- A first application of this topic enables to evaluate the influence of swim fin flexibility on efficiency during swimming based on a CFD structure interaction model. Finite elements simulations are carried out for various material properties and various prescribed kinematics. Besides the significant effect of flexibility on propulsive forces, the results indicate that the propulsive efficiency is greatly influenced by the stroke frequency and the initial angle of attack. For the selected material properties, the results show that efficiency increases from 3.6 percents to 11.9 percents when the stroke frequency is increased from 0 to 1.7 Hz. Moreover efficiency is clearly increased from 5.0 percents to 24.2 percents when increasing the angle of attack from 0 to 45 degrees. Therefore, an interesting prospect of the present work could be an enhancement of the design of better performing swim fins.
- A second application of this topic related to aquatic propulsion deals with a new method to evaluate cross-sectional area based on computer animation of swimming. Indeed, reducing cross sectional area (CSA) during starts and turns is a key part of performance optimisation. Different methods have been used to obtain this parameter without any standard: total human body volume to the power 2/3, wetted area or frontal area based on planimetry technique (PT). These different methods can lead to discrepancies in drag values. Recently, we used two synchronized camcorders to evaluate drag parameters during the different phases of an undulatory stroke cycle. However, such a technique needs accurate synchronization and calibration of the different camcorders views. The aim of this study is to provide a new method based on animation of virtual characters to obtain instantaneous cross-sectional area in an undulatory stroke cycle. Its main advantage is to obtain cross-sectional area as well as biomechanical analysis with a single camcorder in a sagittal plan and without space calibration. A camcorder placed side-on to the swimmer recorded the undulatory movements in the sagittal plane of eight swimmers. This information provided the angles between limbs. These data were then used by our animation engine to animate a virtual swimmer whose anthropometric data came from the real swimmer. A specific algorithm has been developed to automatically obtain the CSA using body outlines. In order to validate our method, we also calculated the CSA using PT with a frontal camcorder view of the same undulatory movements. Our results show similar values of maximum CSA using PT and the frontal camcorder view and our algorithm based on 3D animation. The mean coefficient of variation between the results obtained from the two methods is

7.3 percents. This difference could be related to the level of details of the mesh used to model the avatar. One prospect to this work is to take resistive and propulsive body segments into account in CSA calculation. From this method, we intend to better understand swimming hydrodynamics and the way CSA influences active drag. More generally, this approach has been designed to provide new practical insights into swimming analysis protocols.

6.4. Crowds

Participants: Julien Pettré [contact], Richard Kulpa, Anne-Hélène Olivier, Samuel Lemercier, Jonathan Perrinet, Kevin Jordao.

6.4.1. A realistic model of following behaviors in crowds

Following is an important type of interactions between individuals in crowds. In uni- or bidirectionnal pedestrian traffic, density prevent people from overtaking and going through the crowd: they just start following each other. Based on some experiments performed in the frame of the national project ANR-PEDIGREE, we elaborated a model for simulating following behavior with a very high level of realism. Contributions were presented in [9]. Especially, realism was evaluated both at the microscopic scale and at the macroscopic scale. At the microscopic scale, we carefully reproduce how human do control their motion to follow another walker. At the macroscopic scale, we focused on the emergence of stop-and-go waves that emerge from such traffic. Detailed analysis of experimental data analysis is described in 2 papers in Physical Review E: [15] and [35].

6.5. Interactive Virtual Cinematography

Participants: Marc Christie [contact], Christophe Lino.

The domain of Virtual Cinematography explores the operationalization of rules and conventions pertaining to camera placement, light placement and staging in virtual environments. In 2012, we have tackled two key issues in relation to the reactive control of virtual cameras: (i) the design of an efficient occlusion-free target tracking technique in dynamic environments and (ii) the design of a novel composition technique based on a 2D-manifold representation of search space.

The first issue is related to maintaining the visibility of target objects, a fundamental problem in automatic camera control for 3D graphics applications. Practical real-time camera control algorithms generally only incorporate mechanisms for the evaluation of the visibility of target objects from a single viewpoint, and idealize the geometric complexity of target objects. Drawing on work in soft shadow generation, we perform low resolution projections, from target objects to rapidly compute their visibility for a sample of locations around the current camera position. This computation is extended to aggregate visibility in a temporal window to improve camera stability in the face of partial and sudden onset occlusion. To capture the full spatial extent of target objects we use a stochastic approximation of their surface area. Our implementation is the first practical occlusion-free real-time camera control framework for multiple target objects. The result is a robust component that can be integrated to any virtual camera control system that requires the precise computation of visibility for multiple target (see [20]).

The second challenge is related to the automatic positioning a virtual camera in a 3D environment given the specification of visual properties to be satisfied (on-screen layout of subjects, vantage angles, visibility) is a complex and challenging problem. Most approaches tackle the problem by expressing visual properties as constraints or functions to optimize, and rely on computationally expensive search techniques to explore the solution space. We have shwon how to express and solve the exact on-screen positioning of two or three subjects by expressing the solution space for each couple of subjects as a 2D manifold surface [23]. We demonstrate how to use this manifold surface to solve Blinn's spacecraft problem with a straightforward algebraic approach. We extend the solution to three subjects and we show how to cast the complex 6D optimization problem tackled by most contributions in the field in a simple 2D optimization on the manifold surface by pruning large portions of the search space. The result is a robust and very efficient technique which finds a wide range of applications in virtual camera control and more generally in computer graphics.

We have also explored the application of automated editing techniques to Machinema [19].

Besides we have been involved in the process of rendering camera motions (from real movies) using haptic devices (a joint work with Technicolor and VR4i, accepted at VRST 2012 [21]), and have authored a state of the art report on Haptic Audiovisual (published in Transactions on Haptics [8]).

6.6. Autonomous Virtual Humans

6.6.1. Unifying activity scheduling and path-planning

Participants: Carl-Johan Jorgensen, Fabrice Lamarche [contact].

Crowd distribution in cities highly depends on how people schedule their daily activities. This schedule depends on temporal constraints like appointments or shops opening times. It also relies on the city structure and the locations of the places where activities can be achieved. Personal preferences also affect this schedule: choosing favorite shops or paths for instance.

Within the framework of iSpace&Time project, we are currently working on a model that unifies activity scheduling and path planning into a single process. This process takes city topological configuration into account, as well as time constraints and personal preferences. Applied to thousands of agents, his approach allows us to credibly populate cities. Credible flows of people automatically emerge depending on the time of the day and the city topology.

6.6.2. Long term planning and opportunism

Participants: Philippe Rannou, Fabrice Lamarche [contact].

Autonomous virtual characters evolve in dynamic virtual environments in which changes may be unpredictable. However, they need to behave properly and adapt their behavior to perceived changes while fulfilling their goals. We propose a system that combines long term action planning with failure anticipation and opportunism [27]. The system is based on a modified version of an HTN planning algorithm. It generates plans enriched with information that enable a monitor to detect relevant changes of the environment. Once those changes are detected, a plan adaptation is triggered. Such adaptations include modifying the plan to react to a predicted failure and more importantly to exploit opportunities offered by the environment.

6.6.3. Space-Time planning in dynamic environments

Participants: Thomas Lopez [contact], Fabrice Lamarche [contact].

When automatically populating 3D geometric databases with virtual humanoids, modeling the navigation behavior is essential since navigation is used in most exhibited behaviors. In many application fields, the need to manage navigation in dynamic environments arises (virtual worlds taking physics laws into account, numerical plants in which step stools can be moved,...). This study focuses on the following issue: how to manage the navigation of virtual entities in such dynamic environments where topology may change at any time i.e. where unpredictable accessibility changes can arise at runtime. In opposition to current algorithms, movable items are not only considered as obstacles in the environment but can also help virtual entities in their navigation.

The proposed algorithm [10] splits that problem into two complementary processes: a topology tracking algorithm and a path planning algorithm. The aim of the topology tracking algorithm is to continuously detect and update topological relations between moving objects i.e. accessibility or obstruction, while storing temporal information when recurring relations are observed. The path planning algorithm uses this information to plan a path inside the dynamic environment. The coupling of those algorithms endows a virtual character with the ability to immediately use inserted / moved object to reach previously unreachable locations. Moreover, this algorithm is able to find a path through moving platforms to reach a target located on a surface that is never directly accessible.

MINT Project-Team

6. New Results

6.1. Noisy input filtering for interactive systems

Participants: Géry Casiez [correspondant], Nicolas Roussel.

Noisy signals occur when an original time varying value undergoes undesirable and unpredictable perturbations. These may be caused by things like heat and magnetic fields affecting hardware circuitry, the limits of sensor resolution, or even unstable numerical computation. Noisy signals are a common problem when tracking human motion, particularly with custom sensing hardware and inexpensive input devices like the Kinect or Wiimote.

We developed the 1€ filter ("one Euro filter") is a simple algorithm to filter noisy signals for high precision and responsiveness. It uses a first order low-pass filter with an adaptive cutoff frequency: at low speeds, a low cutoff stabilizes the signal by reducing jitter, but as speed increases, the cutoff is increased to reduce lag. The algorithm is easy to implement, uses very few resources, and with two easily understood parameters, it is easy to tune. When compared with other filters, the 1€ filter shows less lag for a reference amount of jitter reduction [15].

The 1€ filter is already used on a daily basis by many other researchers and companies.

6.2. Transfer functions for subpixel interaction

Participants: Jonathan Aceituno, Géry Casiez [correspondant], Nicolas Roussel.

The current practice of using integer positions for pointing events artificially constrains human precision capabilities (Figure 1). The high sensitivity of current input devices can be harnessed to enable precise direct manipulation "in between" pixels, called subpixel interaction. In [23], we provide a detailed analysis of subpixel theory and implementation, including the critical component of revised control-display gain transfer functions. A prototype implementation is described with several illustrative examples. Guidelines for subpixel domain applicability are provided and an overview of required changes to operating systems and graphical user interface frameworks are discussed.

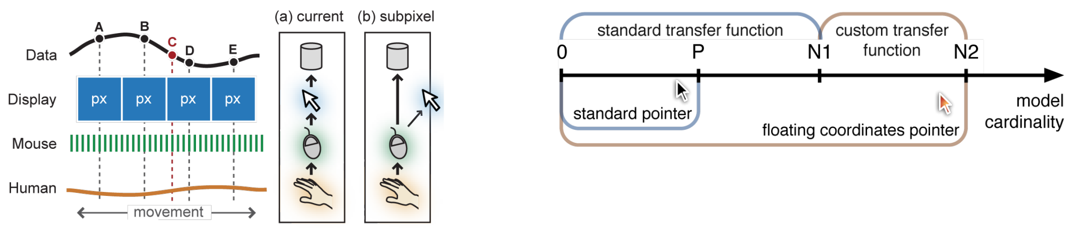
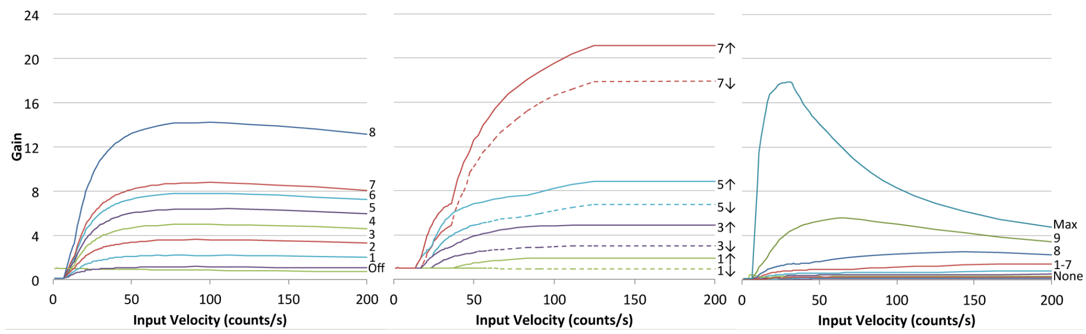


Figure 1. Input mappings: (a) currently, human movements are discretized by mouse sensitivity, then again by display density: data points "in between" pixels like 'C' are unreachable; (b) a subpixel mapping discretizes human movements by mouse sensitivity only, for precise data manipulation (left). Four zones of applicability for subpixel and custom transfer functions (see text for description) (right).

6.3. Transfer functions for scrolling tasks

Participants: Géry Casiez [correspondant], Nicolas Roussel.

Scrolling is controlled through many forms of input devices, such as mouse wheels, trackpad gestures, arrow keys, and joysticks. Performance with these devices can be adjusted by introducing variable transfer functions to alter the range of expressible speed, precision, and sensitivity. However, existing transfer functions are typically "black boxes" bundled into proprietary operating systems and drivers. This presents three problems for researchers: (1) a lack of knowledge about the current state of the field; (2) a difficulty in replicating research that uses scrolling devices; and (3) a potential experimental confound when evaluating scrolling devices and techniques. These three problems are caused by gaps in researchers' knowledge about what device and movement factors are important for scrolling transfer functions, and about how existing devices and drivers use these factors (Figure 2). We fill these knowledge gaps with a framework of transfer function factors for scrolling, and a method for analysing proprietary transfer functions demonstrating how state of the art commercial devices accommodate some of the human control phenomena observed in prior studies [22].



(a) Mac OS X (Generic Device). (b) Microsoft IntelliPoint: up (solid) and down (dashed). (c) Logitech Control Center.

Figure 2. Gain scale factors across input velocity (counts per second) with Mac OS X, Microsoft IntelliPoint (under Windows 7), and Logitech drivers under Mac OS X. Gain is measured as the level of amplification in the system's base unit (pixels per count for Mac OS X and Logitech; lines per count for Microsoft IntelliPoint), and is plotted at varying levels of each driver's respective UI sliders for acceleration.

6.4. Design of transparent tactile stimulators

Participants: Michel Amberg, Frédéric Giraud, Betty Lemaire-Semail [correspondant].

Friction reduction based tactile devices are able to modulate the friction between the fingertip and the active touched surface as a function of fingertip's position. This type of tactile stimulator is thus based on two main components: an active area which vibrates and produces a squeezed air film bearing and a position sensor. Our previous design was made up with a copper plate fully covered by piezo cells, a material which bent when energized by a voltage.

However, this design no longer makes sense when we look forward using tactile feedback on a transparent display. Indeed, for co-localized operation, we can't place piezo cells on the bottom surface of a touch screen since the touched surface would not be transparent; moreover, glass is a non conductive material which complicates the electrical connection.

To cope with these problems, a new design has been introduced. Two copper exciters are firmly bonded on the touch screen to obtain the vibration. These exciters vibrate and propagate their vibration to the glass touch screen. To be efficient, the size of the exciters has to be perfectly adapted to the glass plate. This is why, we not only propose a new way to obtain the vibration of the active area, but we also provide the key design rules of the exciters[19].

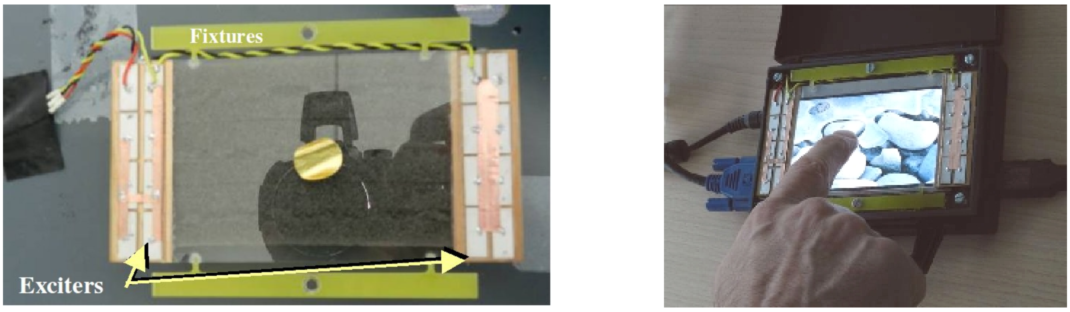


Figure 3. The transparent tactile display, during test procedure (left) and in a co-localized operation (right).

6.5. Methodology for developing textures on friction based interfaces

Participants: Géry Casiez, Thomas Pietrzak, Ludovic Potier, Nicolas Roussel [correspondant], Ibrahim Yapici.

The design of textures for so-called variable friction technologies requires multiple perspectives, which this paper aims to outline and discuss. We first propose a definition of texture and describe the current state of knowledge on their perception. After presenting two technologies for variable friction and comparing them to other tactile interfaces, we describe several particular uses for these devices (Figure 4). We then discuss psychophysical methods for signal perception evaluation and finally discuss methodologies for creating multidimensional tactile content [26].

Figure 4. Examples of textures with increasing complexity in one dimension.

6.6. Hand occlusion on mutitouch surfaces

Participant: Géry Casiez [correspondant].

Operating a computer by directly touching the display surface has many benefits, and in tabletop computing, multi-touch is arguably the most natural form of input. However, with any form of direct input, where the input device and the output display are coincident, the hand and arm cover - or occlude part - of the display. This can be a problem, because compared to manipulating objects on a real tabletop, a tabletop computer is dynamic and can display relevant information, sequential widgets, and system messages in occluded areas.

We examined the shape of hand and forearm occlusion on a multi-touch table for different touch contact types and tasks. Individuals have characteristic occlusion shapes, but with commonalities across tasks, postures, and handedness. Based on this, we create templates for designers to justify occlusion-related decisions and we propose geometric models capturing the shape of occlusion. A model using diffused illumination captures performed well when augmented with a forearm rectangle, as did a modified circle and rectangle model with ellipse "fingers" suitable when only X-Y contact positions are available (Figure 5). Finally, we describe the corpus of detailed multi-touch input data we generated which is available to the community [24].

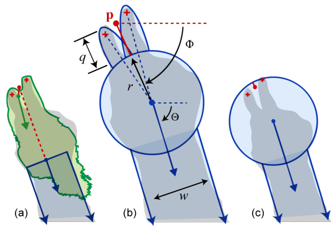


Figure 5. Three occlusion shape models: (a) DI and rectangle; (b) multi-touch circle and rectangle; (c) Vogel et al.

6.7. Indirect multitouch interaction on large screens

Participants: Géry Casiez [correspondant], Jérémie Gilliot, Nicolas Roussel.

Multitouch interaction shows its limits with large display surfaces. Indirect interaction allows to use control surfaces that are much smaller than display surfaces. Absolute indirect interaction raises accuracy problems and relative indirect interaction only allows to interact with a single cursor. We present a relative indirect multitouch interaction technique allowing to create, control, delete several cursors without sacrifying precision for interacting with small objects (Figure 6) [25].

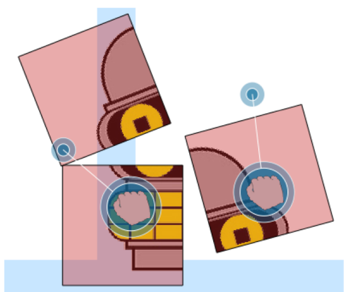


Figure 6. Overview of cursors and cursels used to manipulate two objects.

6.8. Pseudo-rigid movements for flexible multi-finger interactions

Participants: Laurent Grisoni [correspondant], Yosra Rekik, Nicolas Roussel.

Multi-touch interaction requires a trade-off between users' desires and capabilities and gesture recognition constraints. Current approaches to that problem lack flexibility. The number of fingers used for a gesture usually plays a key part in the recognition process, for example. To increase the flexibility of this process, we proposed the use of *pseudo-rigid movements* [27]. We showed how these movements can be determined in real time from the contact information usually available. We explained how they allow to free the recognition process from the number of fingers used and to move towards multi-movement gestures, independent or coordinated. We also presented an interaction technique that takes advantage of this increased flexibility.

6.9. 3D manipulation on multitouch displays

Participants: Anthony Martinet, Géry Casiez [correspondant], Laurent Grisoni.

Multitouch displays represent a promising technology for the display and manipulation of data. While the manipulation of 2D data has been widely explored, 3D manipulation with multitouch displays remains largely unexplored. Based on an analysis of the integration and separation of degrees of freedom, we propose a taxonomy for 3D manipulation techniques with multitouch displays. Using that taxonomy, we introduce Depth-Separated Screen-Space (DS3), a new 3D manipulation technique based on the separation of translation and rotation. In a controlled experiment, we compared DS3 with Sticky Tools and Screen-Space. Results show that separating the control of translation and rotation significantly affects performance for 3D manipulation, with DS3 performing faster than the two other techniques [13].

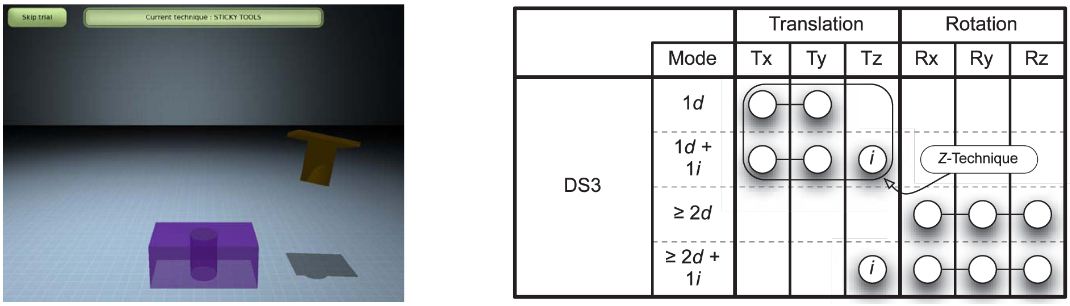


Figure 7. Screen capture of the peg-in-hole task used in the experiment (left). Description of the DS3 technique using the proposed taxonomy (right).

6.10. 3D navigation on multitouch displays

Participants: Clément Moerman, Damien Marchal [correspondant], Nicolas Roussel.

Navigation is one of the elementary tasks of 3d virtual environment. It is composed of two parts: locomotion where there is a physical control of the camera and the wayfinding where a path is found through the environment. Despite being widely studied, there is still need for more efficient and intuitive techniques especially for novice users. Within the context the I-Lab, we worked on a new locomotion technique that combines the advantages of multi-scale navigation and of direct manipulation (Figure 8). The technique, called *Drag'n Go*, was evaluated with a comparative experiment against three other techniques. The results show that *Drag'n Go*: improves performances, reduces learning time and get good user satisfaction either from novice and expert users. The approach and the associated experiment are published in [20].

6.11. Modeling on and above a multitouch surface

Participants: Géry Casiez [correspondant], Bruno De Araujo.

We introduced a semi-immersive environment for conceptual design where virtual mockups are obtained from gestures we aim to get closer to the way people conceive, create and manipulate three- dimensional shapes. We presented on-and-above-the-surface interaction techniques following Guiard's asymmetric bimanual model to take advantage of the continuous interaction space for creating and editing 3D models in a stereoscopic environment. To allow for more expressive interactions, our approach continuously combines hand and finger tracking in the space above the table with multi-touch on its surface. This combination brings forth an alternative design environment where users can seamlessly switch between interacting on the surface or in the space above it depending on the task (Figure 9). Our approach integrates continuous space usage with bimanual interaction to provide an expressive set of 3D modeling operations. Preliminary trials with our experimental setup show this as a very promising avenue for further work [17], [16].

Figure 8. With the Drag'n Go method user can navigate in a 3D virtual environment with a multi-touch device. The movement speed is calculated using perspective based progression scale and it is let under the user's control.

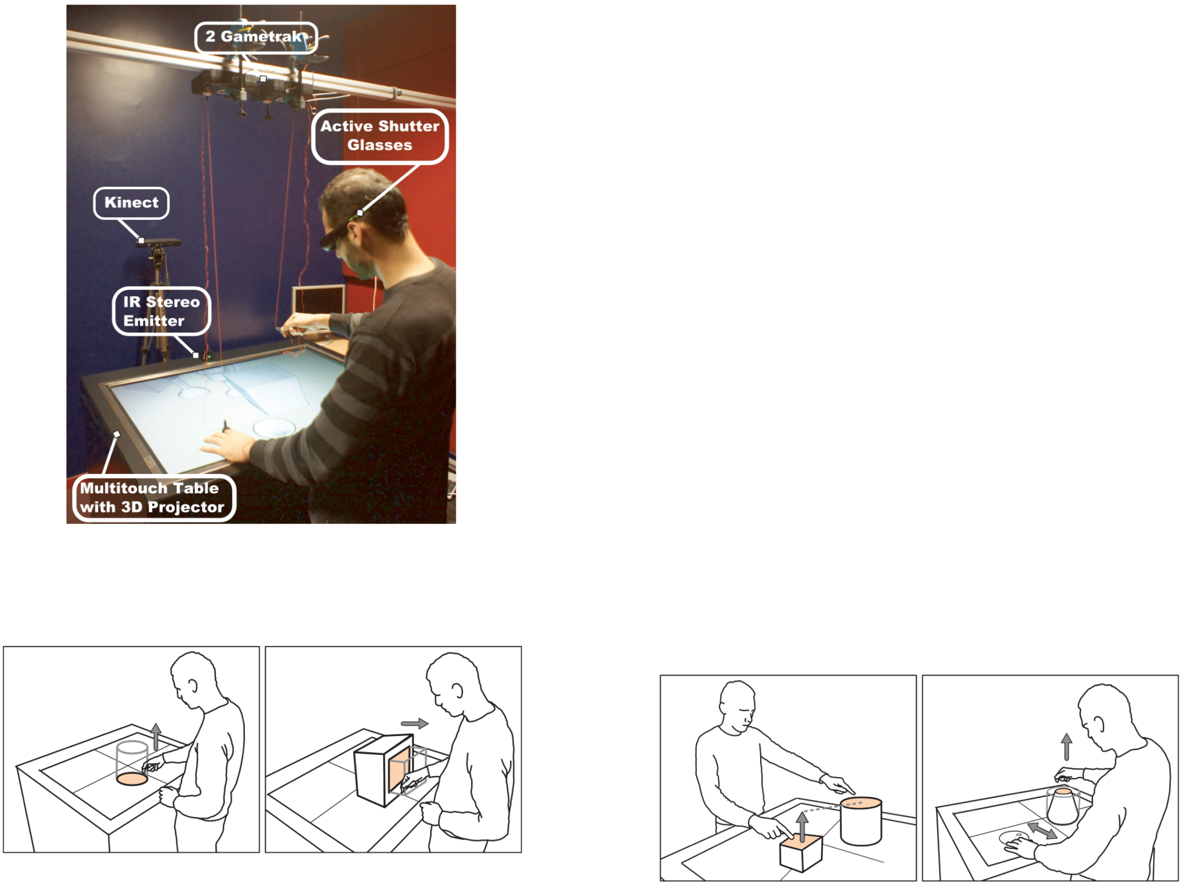


Figure 9. Overview of MockupBuilder setup (left). Examples of face straight extrusion, height constraint and scaling (right).

6.12. Paper-based annotation of digital content from a mobile device

Participant: Thomas Pietrzak [correspondant].

S-Notebook is a hybrid system that makes it possible to take notes on paper about digital content one is exploring on a mobile device (Figure 10). The user can link notes on paper with the content he is currently interacting with, making it possible to reopen it at a later time by tapping the note on his notebook with the digital pen. Therefore he can create bookmarks and hyperlinks on paper notes [21].

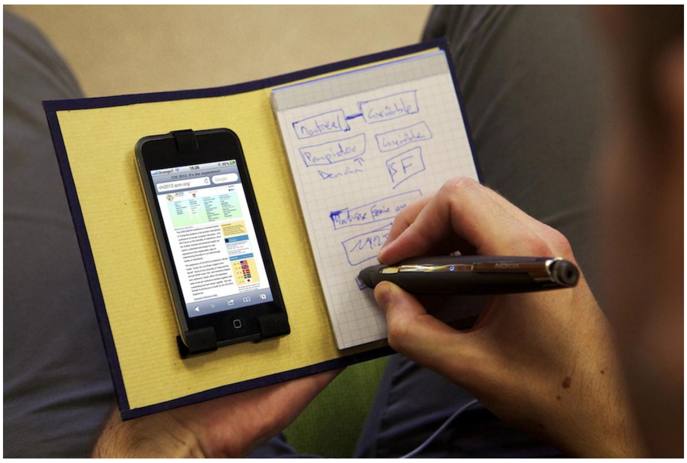


Figure 10. Annotation of digital content on paper.

MORPHEO Team

6. New Results

6.1. A discrete 3D+t Laplacian framework for mesh animation processing

In this work we extend the discrete 3D Laplacian framework to mesh animations, represented as temporally coherent sequences of meshes (Figure 3). In order to let the user control the motion influence with respect to the geometry, we introduce a parameter for the time dimension. Our discrete 3D+t Laplace operator holds the same properties as the discrete 3D Laplacian, as soon as this parameter is non negative. We demonstrate the usefulness of this framework by extending Laplacian-based mesh editing and fairing techniques to mesh animations [15].

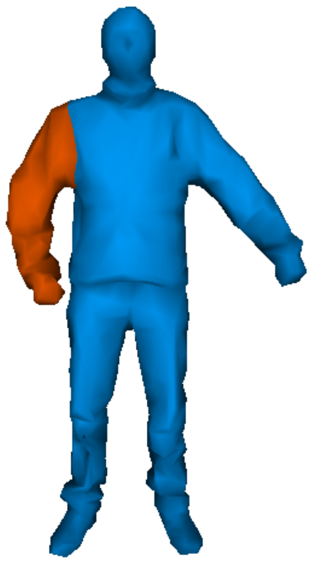


Figure 3. 3D+t Laplacian

6.2. Surface Flow

Recovering dense motion information is a fundamental intermediate step in the image processing chain upon which higher level applications can be built, such as tracking or segmentation. For that purpose, pixel observations in the image provide useful motion cues through temporal variations of the intensity function. We have studied the estimation of dense, instantaneous 3D motion fields over non-rigidly moving surface observed by multi-camera systems. The motivation arises from multi-camera applications that require motion information for arbitrary subjects, in order to perform tasks such as surface tracking or segmentation. To this aim, we have proposed a novel framework that allows to efficiently compute dense 3D displacement fields using low level visual cues and geometric constraints. The main contribution is a unified framework that combines flow constraints for small displacements with temporal feature constraints for large displacements

and fuses them over the surface using local rigidity constraints. The resulting linear optimization problem allows for variational solutions and fast implementations. Experiments conducted on synthetic and real data demonstrated the respective interests of flow and feature constraints as well as their efficiency to provide robust surface motion cues when combined.
As an extension of this work, we also studied the situation where a depth camera and one or more color cameras are available, a common situation with recent composite sensors such as the Kinect. In this case, geometric information from depth maps can be combined with intensity variations in color images in order to estimate smooth and dense 3D motion fields. We propose a unified framework for this purpose, that can handle both arbitrary large motions and sub-pixel displacements. The novelty with respect to existing scene flow approaches is that it takes advantage of the geometric information provided by the depth camera to define a surface domain over which photometric constraints can be consistently integrated in 3D. Experiments on real and synthetic data provide both qualitative and quantitative results that demonstrated the interest of the approach[12].

6.3. Progressive Shape Models

In this work we address the problem of recovering both the topology and the geometry of a deformable shape using temporal mesh sequences (Figure 4). The interest arises in multi-camera applications when unknown natural dynamic scenes are captured. While several approaches allow recovery of shape models from static scenes, few consider dynamic scenes with evolving topology and without prior knowledge. In this nonetheless generic situation, a single time observation is not necessarily enough to infer the correct topology of the observed shape and evidences must be accumulated over time in order to learn this topology and to enable temporally consistent modelling. This appears to be a new problem for which no formal solution exists. We have proposed a principled approach based on the assumption that the observed objects have a fixed topology. Under this assumption, the topology can be progressively learned during the capture of a dynamic scene evolutions. The approach has been successfully experimented on several standard 4D datasets and we believe that it paves the way to more general multi-view scene capture and analysis[8].

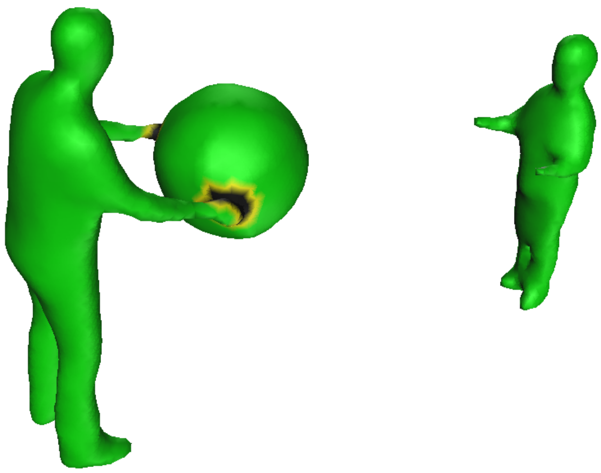


Figure 4. Progressive Shape Models : the balloon can be separated from humans

6.4. Principal Geodesic Dynamics

This work presents a new integration of a data-driven approach using dimension reduction and a physically-based simulation for real-time character animation (Figure 5). We exploit Lie group statistical analysis techniques (Principal Geodesic Analysis, PGA) to approximate the pose manifold of a motion capture sequence by a reduced set of pose geodesics. We integrate this kinematic parametrization into a physically-based animation approach of virtual characters, by using the PGA-reduced parametrization directly as generalized coordinates of a Lagrangian formulation of mechanics. In order to achieve real-time without sacrificing stability, we derive an explicit time integrator by approximating existing variational integrators. Finally, we test our approach in task-space motion control. By formulating both physical simulation and inverse kinematics time stepping schemes as two quadratic programs, we propose a features-based control algorithm that interpolates between the two metrics. This allows for an intuitive trade-off between realistic physical simulation and controllable kinematic manipulation[9].

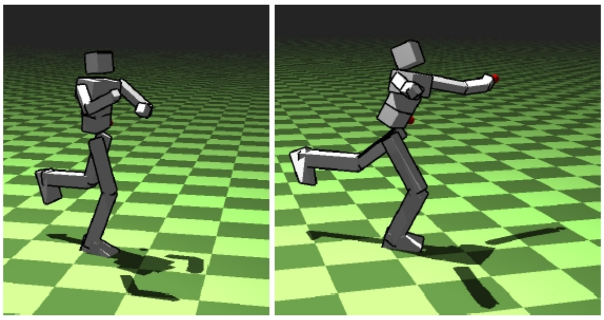


Figure 5. Principal Geodesic Dynamics : test of the balance controller

6.5. A Minimal Solution for Camera Calibration Using Independent Pairwise Correspondences

We have proposed a minimal algorithm for fully calibrating a camera from 11 independent pairwise point correspondences with two other calibrated cameras. Unlike previous approaches, our method neither requires triple correspondences, nor prior knowledge about the viewed scene. This algorithm can be used to insert or re-calibrate a new camera into an existing network, without having to interrupt operation. Its main strength comes from the fact that it is often difficult to find triple correspondences in a camera network. This makes the algorithm, for the specified use cases, probably the most suited calibration solution that does not require a calibration target, and hence can be performed without human interaction [10].

6.6. N-Tuple Color Segmentation for Multi-View Silhouette Extraction

We have presented a new method to extract multiple segmentations of an object viewed by multiple cameras, given only the camera calibration. This method relies on the n-tuple color model to express inter-view consistency when inferring in each view the foreground and background color models permitting the final segmentation. A color n-tuple is a set of pixel colors associated to the n projections of a 3D point. The first goal is set as finding the MAP estimate of background/foreground color models based on an arbitrary sample set of such n-tuples, such that samples are consistently classified, in a soft way, as "empty" if they project in the background of at least one view, or "occupied" if they project to foreground pixels in all views. An Expectation Maximization framework is then used to alternate between color models and soft classifications. In a final step, all views are segmented based on their attached color models. The approach is significantly simpler and faster than previous multi-view segmentation methods, while providing results of equivalent or better quality. [6].

6.7. Cage-based Motion Recovery using Manifold Learning

We have proposed a flexible model-based approach for the recovery of parameterized motion from a sequence of 3D meshes without temporal coherence (Figure 6). Unlike previous model-based approaches using skeletons, we embed the deformation of a reference mesh template within a low polygonal representation of the mesh, namely the cage, using Green Coordinates. The advantage is a less constrained model that more robustly adapts to noisy observations while still providing structured motion information, as required by several applications. The cage is parameterized with a set of 3D features dedicated to the description of human morphology. This allows to formalize a novel representation of 3D meshed and articulated characters, the Oriented Quads Rigging (OQR). To regularize the tracking, the OQR space is subsequently constrained to plausible poses using manifold learning. Results are shown for sequences of meshes, with and without temporal coherence, obtained from multiple view videos preprocessed by visual hull. Motion recovery applications are illustrated with a motion transfer encoding and the extraction of trajectories of anatomical joints. Validation is performed on the HumanEva II database[7].

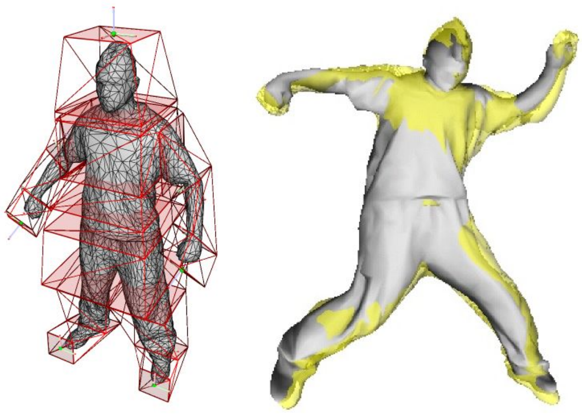


Figure 6. Cage-based Motion Recovery using Manifold Learning

6.8. Segmentation of temporal mesh sequences into rigidly moving components

This work considers the segmentation of meshes into rigid components given temporal sequences of deforming meshes (Figure 7). We have proposed a fully automatic approach that identifies model parts that consistently move rigidly over time. This approach can handle meshes independently reconstructed at each time instant. It allows therefore for sequences of meshes with varying connectivities as well as varying topology. It incrementally adapts, merges and splits segments along a sequence based on the coherence of motion information within each segment. In order to provide tools for the evaluation of the approach, we also introduce new criteria to quantify a mesh segmentation. Results on both synthetic and real data as well as comparisons are provided in the paper[3].

6.9. Keypoints and Local Descriptors of Scalar Functions on 2D Manifolds

This work addresses the problem of describing surfaces using local features and descriptors. While methods for the detection of interest points in images and their description based on local image features are very well

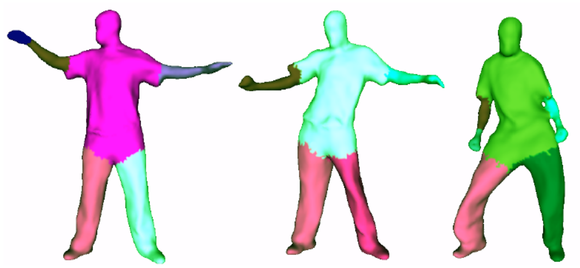


Figure 7. Segmentation of temporal mesh sequences into rigidly moving components

understood, their extension to discrete manifolds has not been well investigated. We provide a methodological framework for analyzing real-valued functions defined over a 2D manifold, embedded in the 3D Euclidean space, e.g., photometric information, local curvature, etc. Our work is motivated by recent advancements in multiple-camera reconstruction and image-based rendering of 3D objects: there is a growing need for describing object surfaces, matching two surfaces, or tracking them over time. Considering polygonal meshes, we propose a new methodological framework for the scale-space representations of scalar functions defined over such meshes. We propose a local feature detector (MeshDOG) and region descriptor (MeshHOG). Unlike the standard image features, the proposed surface features capture both the local geometry of the underlying manifold and the scale-space differential properties of the real-valued function itself. We provide a thorough experimental evaluation. The repeatability of the feature detector and the robustness of feature descriptor are tested, by applying a large number of deformations to the manifold or to the scalar function[4].

MOSTRARE Project-Team

6. New Results

6.1. Modeling XML document transformations

Participants: Joachim Niehren, Angela Bonifati, Sophie Tison, Sławek Staworko, Aurélien Lemay, Anne-Cécile Caron, Yves Roos, Benoît Groz, Antoine Ndione, Tom Sebastian.

XML Schema Validation Groz, Staworko et. al. [26] present a new algorithm that tests determinism of regular expressions in linear time. All regular expressions used in DTDs and XML Schemas are required to be deterministic by the recommendation of the W3C. Whether this is the case can indeed been tested in linear time, as shown in this paper. The best known previous algorithm, which was based on the Glushkov automaton, required O(σ|e|) time, where σ is the number of distinct symbols in e. They also show that matching a word w against a deterministic regular expression e can be achieved in combined linear time O(|e| + |w|) for a wide range of cases.

Staworko et. al. studied bounded repairability for regular tree languages modulo the tree edit distance [28].

Ndione, Niehren, and Lemay [33] present a new probabilistic algorithm for approximate membership of words to regular languages modulo the edit distance on words. In the context of XML, this algorithm is relevant for sublinear DTD validity testing. The time complexity of the algorithm is independent of the size of the input word and polynomial in the size of the input automaton and the inverse error precision. All previous property testing algorithms for regular languages run in exponential time.

XML Query Answering Debarbieux, Niehren, Sebastian et. al. [32] present new algorithms for early XPath node selection on XML Streams. Early selection and rejection is crucial for efficiency, while earliest selection and rejection has high computational complexity in the general case. In contrast to all previous approaches, there algorithm does not rely on any expensive static analysis method. Instead, it is based on a compiler from XPath to nested word automata with selection and rejection states that they introduce. They cover a large fragment of downward XPath, with the main restriction that negation is forbidden above descendant axis and disjunctions. Non-determinism is used to deal with descendant axis and disjuctions. High run-time efficiency in practice is obtained by on-the-fly determinization for nested word automata, even in cases where static determinization produces automata of more than exponential size. Our experimental results confirm a very high efficiency in space and time. An implementation of our FXP/QuiXPath system is freely available and used for industrial transfer in the QuiXProc system.

Staworko et. al. tackled prioritized repairing and consistent query answering in relational databases in [20].

External Cooperations with other teams in Lille lead to the following publications [19], [31], [30].

6.2. Machine learning for XML document transformations

Participants: Adrien Boiret, Jean Decoster, Pascal Denis, Jean-Baptiste Faddoul, Antonino Freno, Gemma Garriga, Rémi Gilleron, Mikaela Keller, Grégoire Laurence, Aurélien Lemay, Joachim Niehren, Sławek Staworko, Marc Tommasi, Fabien Torre.

Learning XML Queries. Staworko et. al. [29] proposed learning twig and path queries.

Niehren, Champavère, Gilleron, and Lemay [34] propose new algorithm and learnability result for XML query induction based on schema-guided pruning strategies. Pruning strategies impose additional assumptions on node selection queries that are needed to compensate for small numbers of annotated examples. The class of regular queries that are stable under a given schema-guided pruning strategy was distinguished and shown to be learnable with polynomial time and data. The learning algorithm is obtained by adding pruning heuristics to the traditional learning algorithm for tree automata from positive and negative examples. While justified by a formal learning model, their learning algorithm for stable queries also performs very well in practice of XML information extraction.

Learning XML Transformations. Boiret, Lemay, and Niehren [21] solved the long open question of how to learn rational functions with polynomial time and data. Rational functions are transformations from words to words that can be defined by deterministic string transducers with lookahead. No previous learning results for classes of transducers with look-ahead existed, so this results is relevant for learning XML transformations defined by transducers with look-ahead, as with XSLT.

Multi-task Learning. We address the problem of multi-task learning with no label correspondence among tasks. In [22], Faddoul, Chidlovskii, Gilleron and Torre propose the multi-task Adaboost algorithm with Multi-Task Decision Trees as weak classifiers. They conduct experiments on multi-task datasets, including the Enron email set and Spam Filtering collection. Faddoul successfully defended his PhD thesis [16] in June 2012.

Probabilistic models for large graphs. We propose new approaches for the statistical analysis of large-scale undirected graphs. The guiding idea is to exploit the spectral decomposition of subgraph samples, and in particular their Fiedler eigenvalues, as basic features for density estimation and probabilistic inference. In [24], Freno, Keller, Garriga, and Tommasi develop a conditional random graph model for learning to predict links in information networks (such as scientific coauthorship and email communication). In [25], Freno, Keller, and Tommasi propose instead to estimate joint probability distributions through (non-linear) random fields, applying the resulting model to graph generation and link prediction.

Learning in Multiple graphs Ricatte, Garriga, Gilleron and Tommasi focus on learning from several sources of heterogeneous data. They represent each source as a graph of data and they propose to combine the multiple graphs with the help of small number of labeled nodes. They obtain a kernel that can be used as input to different graph-learning tasks such as node classification and clustering. The paper is under submission. Along a collaboration with physicians, Keller and Tommasi consider graphs that represents the structural connectivity of the brain (connectome). They develop a spatially constrained clustering method, combining heterogeous descriptions of the same objects through the graph of neighborhood on the cortex and the graph of connectivity. The paper is under submission.

Starting PhDs Boneva, Bonifati and Staworko started to supervise the PhD of R. Ciucanu on learning cross-model database mappings. Denis and Tommasi has begun to supervise the PhD of David Chatel on guided clustering for graphs (of texts).

OAK Team

6. New Results

6.1. Efficient XML and RDF data management

6.1.1. Efficient and safe management of XML and JSON data

We addressed the problem of detecting independence between XML queries and updates. Since the problem is undecidable for XQuery queries and updates, and is intractable even for restricted fragments, we adopted an approximating technique based on a schema-based static analysis. Our analysis turned to be precise and, at the same time, fast to run. Main result about this research line have been published in [6], while the complete study is reported in Federico Ulliana's PhD Thesis (defended in December 12) [5].

To address the problem of manipulating large XML documents via main-memory XQuery engines, largely used for their efficiency and easiness of integration in a programming environment, we developed partitioning techniques for both XQuery queries and updates. Our technique is based on a static analysis over queries and updates (no schema is used) able to infer information that is used to partition the input document, in a streaming fashion. Besides allowing existing main-memory system to scale up in terms of query/update input size, our technique also admits a MapReduce implementation. Main results have been published in [11], while the complete study is reported in Noor Malla's PhD Thesis (defended on September 21) [3].

We also tackled the problem of safe manipulation of JSON data. Some typed and MapReduce-based programming languages for manipulating JSON data have been recently proposed. However, the problem of inferring a schema for untyped JSON data was still open, and having a schema for manipulated data is fundamental for the afore mentioned programming languages. We started investigating technique able to deal with massive JSON data sets. To ensure efficiency, our technique is based on Map-Reduce, while to ensure precision and conciseness it adopts type rewriting rules able to: i) compact as much as possible intermediate inferred types, and ii) to avoid gross approximation when compacting types. Some preliminary results are quite encouraging, and appeared in [21].

6.1.2. Hybrid models for XML and RDF

Considerable energy is spent towards enriching XML data on the web with semantics through annotations. These annotations can range from simple metadata to complex semantic relationships between data items. Although the vision of supporting such annotations is spreading, it still lacks the infrastructure that will enable it. To this end we have proposed a framework enabling the storage and querying of annotated documents. We have introduced (i) the XR data model, in which annotated documents are XML documents described by RDF triples and (ii) the query language XRQ to interrogate annotated documents through their structure and their semantics. A prototype platform XRP for the management of annotated documents has also been developed, to show the relevance of our approach through experiments [9].

6.1.3. RDF query answering

A promising method for efficiently querying RDF data consists of translating SPARQL queries into efficient RDBMS-style operations. However, answering SPARQL queries requires handling *RDF reasoning*, which must be implemented outside the relational engines that do not support it. We have introduced the *database (DB) fragment of RDF*, going beyond the expressive power of previously studied RDF fragments. Within this fragment, we have devised novel *sound and complete techniques* for answering *Basic Graph Pattern (BGP)* queries, exploring the two established approaches for handling RDF semantics, namely reformulation and saturation. In particular, we have focused on handling database *updates* within each approach and proposed a method for incrementally maintaining the saturation; updates raise specific difficulties due to the rich RDF semantics. Our techniques have been designed to be deployed on top of any RDBMS(-style) engine, and we have experimentally studied their performance trade-offs [20], [14], [25].

6.1.4. Efficient and scalable Web Data Entity Resolution

We addressed the problem of detecting multiple heterogeneous representations of a real-world object (often referred to as record linkage, duplicate detection, or entity resolution) in two contexts, i.e., for hierarchical data and for data where relationships between entities form a graph.

Concerning XML entity resolution, we contributed to a novel algorithm that uses a Bayesian network to determine the probability of two XML elements being duplicates. The probability is based both on content and on structure information given by the hierarchical XML model. To efficiently evaluate the Bayesian network to find duplicates, we devised two pruning techniques. Whereas the first is lossless in terms of not loosing any true duplicates, the second pruning heuristic trades off runtime for a somewhat lower accuracy of the duplicate detection result. An experimental evaluation shows that the proposed solutions are capable of outperforming other state-of-the art XML duplicate detection methods [8].

As for duplicate detection in entity graphs, we defined a general framework for algorithms tackling this problem. The general process consists of three steps, namely retrieval, classification, and update. We further proposed an algorithm complying to the framework that leverages an off-the-shelf relational database to store and to efficiently query information (both data and relationships) relevant for duplicate classification. We further extended our framework and algorithm to allow for parallel and batched processing. Our experimental validation on data of up to two orders of magnitude larger than data considered by other state-of-the-art algorithms showed that the proposed methods allow to scale duplicate detection in entity graphs to large volumes of data [7].

6.1.5. Warehousing RDF data

Data warehousing (DW) research has lead to a set of tools and techniques for efficiently analyzing large amounts of multi-dimensional data. As more data gets produced and shared in RDF, analytic concepts and tools for analyzing such irregular, graph-shaped, semantic-rich data are needed. We have introduced *the first all-RDF model for warehousing RDF graphs*. Notably, we have defined *RDF analytical schemas*, themselves full RDF graphs, and *RDF analytical queries*, corresponding to the relational DW star/snowflake schemas and cubes. We have shown how *RDF OLAP operations* can be performed on our RDF cubes. We have also performed experiments validating the practical interest of our approach.

6.2. Cloud-based Data Management

We investigate architectures for storing Web data (in particular, XML documents and RDF graphs) based on commercial cloud platforms. In particular, we have developed the AMADA platform, which operates in a Software as a Service (SaaS) approach, allowing users to upload, index, store, and query large volumes of Web data. Since cloud users support monetary costs directly connected to their consumption of cloud resources, we focus on indexing content in the cloud. We study the applicability of several indexing strategies, and show that they lead not only to reducing query evaluation time, but also, importantly, to reducing the monetary costs associated with the exploitation of the cloud-based warehouse [10], [12], [13].

6.3. Data Transformation Management

When developing data transformations – a task omnipresent in applications like data integration, data migration, data cleaning, or scientific data processing – developers quickly face the need to verify the semantic correctness of the transformation. Declarative specifications of data transformations, e.g. SQL or ETL tools, increase developer productivity but usually provide limited or no means for inspection or debugging. In this situation, developers today have no choice but to manually analyze the transformation and, in case of an error, to (repeatedly) fix and test the transformation.

The above observations call for a more systematic management of a data transformation. Within Oak, we have so far focused on the first phase of the process described above, namely the analysis phase. Leveraging results obtained in previous years (by us and others), we solidified the theory of why-not provenance. Analogously to a distinction between different types of why-provenance, we defined three types of why-not provenance. For each of the three types, we surveyed the semantics employed by different approaches, e.g., set vs. bag semantics or existential vs. universal quantification. We also identified cases of implication and equivalence between why-not provenance of different types. We have leveraged this theoretical work during the design of a novel algorithm that has the potential to overcome usability and efficiency limitations of previous algorithms after further optimization, implementation, and validation in the future. Furthermore, we implemented different approaches for why-provenance and why-not provenance and included them in the Nautilus Analyzer, a system prototype for declarative query debugging. We demonstrated this prototype at CIKM 2012 [15].

ORPAILLEUR Project-Team

6. New Results

6.1. The Mining of Complex Data

Participants: Mehwish Alam, Thomas Bourquard, Aleksey Buzmakov, Victor Codocedo, Adrien Coulet, Elias Egho, Nicolas Jay, Florence Le Ber, Ioanna Lykourentzou, Luis Felipe Melo, Amedeo Napoli, Chedy Raïssi, My Thao Tang, Yannick Toussaint.

formal concept analysis, relational concept analysis, pattern structures, search for frequent itemsets, association rule extraction, mining of complex data, graph mining, skylines, sequence mining, FCA in spatial and temporal reasoning

Formal concept analysis, together with itemset search and association rule extraction, are suitable symbolic methods for KDDK, that may be used for real-sized applications. Global improvements may be carried on the scope of applicability, the ease of use, the efficiency of the methods, and on the ability to fit evolving situations. Accordingly, the team is working on extensions of such symbolic data mining methods to be applied on complex data such as biological or chemical data or textual documents, involving objects with multi-valued attributes (e.g. domains or intervals), n-ary relations, sequences, trees and graphs.

6.1.1. FCA, RCA, and Pattern Structures

Recent advances in data and knowledge engineering have emphasized the need for Formal Concept Analysis (FCA) tools taking into account structured data. There are a few extensions of FCA for handling contexts involving complex data formats, e.g. graphs or relational data. Among them, Relational Concept Analysis (RCA) is a process for analyzing objects described both by binary and relational attributes [116]. The RCA process takes as input a collection of contexts and of inter-context relations, and yields a set of lattices, one per context, whose concepts are linked by relations. RCA has an important role in KDDK, especially in text mining [86], [85].

Another extension of FCA is based on Pattern Structures (PS) [94], which allows to build a concept lattice from complex data, e.g. nominal, numerical, and interval data. In [101]), pattern structures are used for building a concept lattice from intervals, in full compliance with FCA, thus benefiting of the efficiency of FCA algorithms. Actually, the notion of similarity between objects is closely related to these extensions of FCA: two objects are similar as soon as they share the same attributes (binary case) or attributes with similar values or the same description (at least in part). Various results were obtained in the study of the relations existing between FCA with an embedded explicit similarity measure and FCA with pattern structures [100]. Moreover, similarity is not a transitive relation and this lead us to the study of tolerance relations. In addition, a new research perspective is aimed at using frequent itemset search methods for mining interval-based data being guided by pattern structures and biclustering as well.

6.1.2. Advances in FCA and Pattern Mining

In the context of environmental sciences, research work is in concern with the mining of complex hydroecological data with concept lattices. In particular, Florence Le Ber –as a member of UMR 7517 Lhyges, Strasbourg– is the scientific head of an ANR project named "FRESQUEAU" (2011–2014) dealing with FCA and data mining and hydroecological data (see http://engees.unistra.fr/site/recherche/projets/anr-fresqueau/).

In this framework, concept lattices based on multi-valued contexts have been used for characterizing macroinvertebrate communities in wetland and their seasonal evolution [19]. Within the ANR Fresqueau project we are studying tools for sequential pattern extraction taking into account spatial relations [56], [43].

From another point of view, miscanthus is a perennial crop used for biomass production. Its implantation is rather new, and there is few farms cultivating miscanthus in France. Understanding the farmers' choices for allocating miscanthus in their farmland is a main challenge. The CBR model is investigated for modeling these choices from farm surveys, including spatial reasoning aspects [20], [47] [41].

For completing the work on FCA and itemset search, there is still on-going work on frequent and rare itemset search, for being able to build lattices from very large data and completing the algorithm collection of the Coron platform. Work is still in progress on the design of an integrated and modular algorithm for searching for closed and generators itemsets, and equivalence classes of itemsets, thus enabling the construction of the associated lattice [121]. This research aspect is also linked to the research carried on within a the PICS CaDoE research project (see Section 8.1.1.3). In addition, there is also research work carried on different aspects involving the management of big data in the context of the BioIntelligence Project and the Quaero Project.

6.1.3. Skylines, sequential data, privacy and E-sports analytics

Pattern discovery is at the core of numerous data mining tasks. Although many methods focus on efficiency in pattern mining, they still suffer from the problem of choosing a threshold that influences the final extraction result. One goal is to make the results of pattern mining useful from a user-preference point of view. That is, take into account some domain knowledge to guide the pattern mining process. To this end, we integrate into the pattern discovery process the idea of skyline queries in order to mine *skyline patterns* in a threshold-free manner. This forms the basis for a novel approach to mining skyline patterns. The efficiency of our approach was illustrated over a use case from *chemoinformatics* and we showed that small sets of dominant patterns are produced under various measures that are interesting for chemical engineers and researchers.

Sequence data is widely used in many applications. Consequently, mining sequential patterns and other types of knowledge from sequence data has become an important data mining task. The main emphasis has been on developing efficient mining algorithms and effective pattern representation.

However, important fundamental problems still remained open: (i) given a sequence database, can we have an upper bound on the number of sequential patterns in the database? (ii) Is the efficiency of the sequence classifier only based on accuracy? (iii) Do the classifiers need the entire set of extracted patterns or a smaller set with the same expressiveness power?

In the field of the management of sequential date in medicine, analysis of health care trajectories led to the development of a new sequential pattern mining method [42]. The MMISP algorithm is able to efficiently extract sequential patterns composed of itemsets and multidimensional items. The multidimensional items can be described with additional taxonomic knowledge, allowing mining with appropriate levels of granularity. In parallel, a new measure has been created to compute the similarity between sequences of itemsets [78].

Orpailleur is one of the few project-teams working on privacy challenges which are becoming a core issue with different scientific problems in computer science. With technology infiltrating more and more every aspect of our lives, each human activity leaves a digital trace in some repository. Vast amounts of personal data are implicitly or explicitly created each day, and rarely one is aware of the extent of information that is kept, processed and analyzed without his knowledge or consent. These personal data give rise to significant concerns about user privacy, since important and sensitive details about private life are collected and exploited by third parties. The goal of privacy preservation technologies is to provide tools that allow greater control over the dissemination of user data. A promising trend in the field is Privacy Preserving Data Publishing (PPDP), which allows sharing of anonymized data. Anonymizing a dataset is not limited to the removal of direct identifiers that might exist in a dataset, e.g. the full name or the Social Security Number of a person. It also includes removing secondary information, e.g. like age, zip code that might lead indirectly to the true identity of an individual.

Existing research on this problem either perturbs the data, publishes them in disjoint groups disassociated from their sensitive labels, or generalizes their values by assuming the availability of a generalization hierarchy. In a recent work, we proposed a novel alternative [54]. Our publication method also puts data in a generalized form, but does not require that published records form disjoint groups and does not assume a hierarchy either. Instead, it employs generalized bitmaps and recasts data values in a nonreciprocal manner.

One of the most fascinating challenges of our time is understanding the complexity of the global interconnected society we inhabit. Today we have the opportunity to observe and measure how our society intimately works, by analyzing the big data. i.e, the digital breadcrumbs of human activities sensed as a by-product of the ICT systems that we use. These data describe the daily human activities: for instance, automated payment systems record the tracks of our purchases, search engines record the logs of our queries for finding information on the web, social networking services record our connections to friends, colleagues and collaborators, wireless networks and mobile devices record the traces of our movements and our communications. These social data are at the heart of the idea of a knowledge society, where decisions can be taken on the basis of knowledge in these data.

Social network data analysis raises concerns about the privacy of related entities or individuals. We theoretically establish that any kind of structural identification attack can effectively be prevented using random edge perturbation and show that, surprisingly, important properties of the whole network, as well as of subgraphs thereof, can be accurately calculated and hence data analysis tasks performed on the perturbed data, given that the legitimate data recipient knows the perturbation probability as well [53].

"Electronic-sport" (E-Sport) is now established as a new entertainment genre. More and more players enjoy streaming their games, which attract even more viewers. In fact, in a recent social study, casual players were found to prefer watching professional gamers rather than playing the game themselves. Within this context, advertising provides a significant source of revenue to the professional players, the casters (displaying other people's games) and the game streaming platforms. In a recent work with Mehdi Kaytoue, we started focusing on the huge amount of data generated by electronic games. We crawled, during more than 100 days, the most popular among such specialized platforms: Twitch.tv.

Thanks to these gigabytes of data, we proposed a first characterization of a new Web community, and we showed, among other results, that the number of viewers of a streaming session evolves in a predictable way, that audience peaks of a game are explainable and that a Condorcet method can be used to sensibly rank the streamers by popularity [45]. This work should bring to light the study of E-Sport and its growing community for computer scientists and sociologists. They indeed deserve the attention of industrial partners (for the large amount of money involved) and researchers (for interesting problems in social network dynamics, personalized recommendation, sentiment analysis, etc.).

6.1.4. KDDK in Text Mining

Ontologies help software and human agents to communicate by providing shared and common domain knowledge, and by supporting various tasks, e.g. problem-solving and information retrieval. In practice, building an ontology depends on a number of “ontological resources” having different types: thesaurus, dictionaries, texts, databases, and ontologies themselves. We are currently working on the design of a methodology and the implementation of a system for ontology engineering from heterogeneous ontological resources. This methodology is based on both FCA and RCA, and was previously successfully applied in contexts such as astronomy and biology. At present, an engineer is implementing a robust system being guided by the previous research results and preparing the way for some new research directions involving trees and graphs (see also the work on the ANR Hybride project).

6.2. KDDK in Life Sciences

Participants: Yasmine Assess, Emmanuel Bresso, Thomas Bourquard, Adrien Coulet, Marie-Dominique Devignes, Anisah Ghoorah, Renaud Grisoni, Jean-François Kneib, Florence Le Ber, Bernard Maigret, Jean-François Mari, Amedeo Napoli, Violeta Pérez-Nueno, Dave Ritchie, Malika Smaïl-Tabbone.

The Life Sciences constitute a challenging domain in which to implement knowledge-guided approaches for knowledge discovery. Biological data are complex from many points of views: voluminous, high-dimensional, deeply inter-connected, etc. Analyzing such data and extracting hidden knowledge has become a crucial issue in important domains such as health, environment and agronomy. More and more bio-ontologies are available and can be used to enhance the knowledge discovery process [88], [117]. In the next few years, the experience of the Orpailleur team in KDDK applied to the Life Sciences will be further developed in two directions: the use of bio-ontologies to improve approaches for data integration and mining when applied to real-world data, and the study of the synergy between numeric and symbolic data-mining methods in life-science applications.

6.2.1. Relational data mining applied to complex biological object characterization and prediction

Inductive Logic Programming (ILP) is a learning method which allows expressive representation of the data and produces explicit first-order logic rules. However, any ILP system returns a single theory based on heuristic user-choices of various parameters and learning biases, thus ignoring potentially relevant rules. Accordingly, we propose an approach based on Formal Concept Analysis for effective interpretation of reached theories with the possibility of adding domain knowledge. Our approach was applied to the characterization of three-dimensional (3D) protein-binding sites, namely phosphorylation sites, which are the protein portions on which interactions with other proteins take place [33]. In this context, we defined a logical representation of 3D patches and formalized the problem as a concept learning problem using ILP. Another application of this KDDK methodology concerns the characterization and prediction of drug side-effect profiles (Journal manuscript in preparation). In this case, maximal frequent itemsets are extracted and allow us to propose relevant side-effect profiles of drugs which are further characterized by ILP.

6.2.2. Functional classification of genes using semantic similarity matrix and various clustering approaches

In the last report, we proposed a measure called IntelliGO which computes semantic similarity between genes for discovering biological functions shared by a set of genes (e.g., showing the same expression profile). This measure takes into account domain knowledge represented in Gene Ontology (GO) [83].

Functional classification aims at grouping genes according to their molecular function or the biological process they participate in. Evaluating the validity of such unsupervised gene classification remains a challenge given the variety of distance measures and classification algorithms that can be used. We evaluated functional classification of genes with the help of reference sets. Overlaps between clusters and reference sets are estimated by the F-score metric. We test the IntelliGO measure with hierarchical and fuzzy C-means clustering algorithms and we compare results with the state-of-the-art DAVID functional classification method (Database for Annotation Visualization and Integrated Discovery). Finally, study of best matching clusters to reference sets leads us to propose a method based on set-differences for discovering missing information.

The IntelliGO-based functional clustering method was tested on four benchmarking datasets consisting of biological pathways (KEGG database) and functional domains (Pfam database) [13]. The IntelliGO measure is usable on line (see http://bioinfo.loria.fr/Members/benabdsi/intelligo_project/).

We are currently investigating the clustering problem when objects are not represented as feature vectors in a vector space but as a pairwise similarity matrix. In biology such similarity measures are often computationally expensive or incompatible with *bona fide* distance definition. Embedding techniques of pairwise data into Euclidean space aim at facilitating subsequent clustering of the objects [115]. Spectral clustering methods are also relevant in this case [127]. We are conducting comparative and large-scale gene clustering evaluation using the Intelligo measure and reference sets.

6.2.3. Analysis of biomedical data annotated with ontologies

Annotating data with concepts of an ontology is a common practice in the biomedical domain. Resulting annotations define links between data and ontologies that are key for data exchange, data integration and data analysis tasks. In 2011 we collaborated with the National Center for Biomedical Ontologies (NCBO) to develop of large repository of annotations named the NCBO Resource Index [99]. The resulting repository contains annotations of 34 biomedical databases annotated with concepts of 280 ontologies of the BioPortal [2]. We proposed a comparison of the annotations of a database of biomedical publications (Medline) with two databases of scientific funding (Crisp and ResearchCrossroads) to profile disease research [18]. The annotation of these three databases with a unique ontology about diseases enable to consider their content conjointly and consequently to analyze and compare, for distinct disease (or family of diseases), trends in term of number of publications and funding amounts.

We started a new project that aims at exploring biomedical annotations with FCA techniques. One main challenge here is to develop a knowledge discovery approach that consider the knowledge represented in the ontologies employed for the annotations.

6.2.4. Connecting textual biomedical knowledge with the Semantic Web

A large amount of biomedical knowledge is in the form of text embedded in published articles, clinical files or biomedical public databases. It is consequently of high interest to extract and structure this knowledge to facilitate its consideration when processing biomedical data. We benefited from advances in Natural Language Processing (NLP) techniques to extract fine-grained relationships mentioned in biomedical text and subsequently published such relationships on line in the form of RDF triples [91], [90]. In a collaborative work with the Health Care and Life Science (HCLS) interest group of the W3C, we demonstrated how biomedical knowledge extracted from text, along with Semantic Web technologies has high potential for recommendation systems and knowledge discovery in biomedicine [118].

6.3. Structural Systems Biology

Participants: Thomas Bourquard, Marie-Dominique Devignes, Anisah Ghoorah, Van-Thai Hoang, Bernard Maigret, Violeta Pérez-Nueno, Dave Ritchie, Malika Smaïl-Tabbone.

knowledge discovery in life sciences, bioinformatics, biology, chemistry, gene

Structural systems biology aims to describe and analyze the many components and interactions within living cells in terms of their three-dimensional (3D) molecular structures. We are currently developing advanced computing techniques for molecular shape representation, protein-protein docking, protein-ligand docking, high-throughput virtual drug screening, and knowledge discovery in databases dedicated to protein-protein interactions.

6.3.1. Accelerating protein docking calculations using graphics processors

We have recently adapted the *Hex* protein docking software [113] to use modern graphics processors (GPUs) to carry out the expensive FFT part of a docking calculation [114]. Compared to using a single conventional central processor (CPU), a high-end GPU gives a speed-up of 45 or more. This software is publicly available at http://hex.loria.fr. A public GPU-powered server has also been created (http://hexserver.loria.fr) [105]. The docking server has performed some 12,000 docking runs during 2012. A book chapter describing the Hex docking algorithm has been published [75]. Our docking work has facilitated further developments on modeling the assembly of multi-component molecular structures using a particle swarm optimization technique [25], and on modeling protein flexibility during docking [24].

6.3.2. KBDOCK: Protein docking using Knowledge-Based approaches

In order to explore the possibilities of using structural knowledge of protein-protein interactions, Anisah Ghoorah recently developed the KBDOCK system as part of her doctoral thesis project. KBDOCK combines residue contact information from the 3DID database [119] with the Pfam protein domain family classification [92] together with coordinate data from the Protein Data Bank [87] in order to describe and analyze all known protein-protein interactions for which the 3D structures are available. We have demonstrated the utility of KBDOCK [96] for template-based docking using 73 complexes from the Protein Docking Benchmark [98]. KBDOCK is available at http://kbdock.loria.fr. Anisah Ghoorah successfully defended her thesis in November 2012 [10].

[2] http://bioportal.bioontology.org/

6.3.3. Kpax: A new algorithm for protein structure alignment

We have developed a new protein structure alignment approach called Kpax [6]. The approach exploits the fact that each amino acid residue has a carbon atom with a highly predictable tetrahedral geometry. This allows the local environment of each residue to be transformed into a canonical orientation, thus allowing easy comparison between the canonical orientations of residues within pairs of proteins using a novel scoring function based on Gaussian overlaps. The overall approach is two or three orders of magnitude faster than most contemporary protein structure alignment algorithms, while still being almost as accurate as the state-of-the-art TM-Align approach [126]. The Kpax program is available at http://kpax.loria.fr/.

6.3.4. gEMpicker and gEMfitter: GPU-accelerated tools for cryo-electron microscopy

Solving the structures of large protein assemblies is a difficult and computationally intensive task. Multiple two-dimensional (2D) images must be processed and classified to identify protein particles in different orientations. These images may then be averaged and stacked to deduce the three-dimensional (3D) structure of a protein. In order to help accelerate the first of these tasks we have recently developed a novel and highly parallel algorithm called “gEMpicker” which uses multiple graphics processors to detecting 2D particles in cryo-electron microscopy images. We have also developed a 3D shape matching algorithm called “gEMfitter” which also exploits graphics processors, and which will provide a useful tool for the final 3D assembly step. Both programs will soon be made publicly available, and two manuscripts describing our approach are in preparation.

6.3.5. DOVSA: Developing new algorithms for virtual screening

In 2010, Violeta Pérez-Nueno joined the Orpailleur team thanks to a Marie Curie Intra-European Fellowship (IEF) award to develop new virtual screening algorithms (DOVSA). The aim of this project is to advance the state of the art in computational virtual drug screening by developing a novel consensus shape clustering approach based on spherical harmonic (SH) shape representations [111]. The main disease target in this project is the acquired immune deficiency syndrome (AIDS), caused by the human immuno-deficiency virus (HIV) [109]. However, the approach will be quite generic and will be broadly applicable to many other diseases. Good progress has been made on calculating and clustering spherical harmonic “consensus shapes” which represent rather well the essential features of groups of active molecules [110]. The approach has since been extended to provide a rapid way to cluster drug families according to the Gaussian distributions of their surface shapes, and to predict possible cross-interactions of drug families [21]. We have also published a review on the state of the art in 3D virtual drug screening [15].

6.4. Around the Taaable research project

Participants: Valmi Dufour-Lussier, Emmanuelle Gaillard, Laura Infante Blanco, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer.

knowledge representation, description logics, classification-based reasoning, case-based reasoning, belief revision, semantic web

The Taaable project (http://taaable.fr) has been originally created as a challenger of the Computer Cooking Contest (ICCBR Conference). A candidate to this contest is a system whose goal is to solve cooking problems on the basis of a recipe book (common to all candidates), where each recipe is a shallow XML document with an important plain text part. The size of the recipe book (about 1500 recipes) prevents from a manual indexing of recipes: this indexing is performed using semi-automatic techniques.

Beyond its participation to the CCCs, the Taaable project aims at federating various research themes: case-based reasoning, information retrieval, knowledge acquisition and extraction, knowledge representation, minimal change theory, ontology engineering, semantic wikis, text-mining, etc. Case-based reasoning is used to perform adaptation of recipe to user constraints. The reasoning process uses a cooking domain ontology (especially hierarchies of classes) and adaptation rules. The knowledge base used by the inference engine is encoded within a semantic wiki, which contains the recipes, the domain ontology, and adaptation rules.

The most important original features of this version are:

Modules for computing adaptation knowledge. Using adaptation knowledge, and especially adaptation rules, is a way to better adapt cooking recipes to user constraints. A previous work for extracting adaptation rules has been performed in 2011 [93]. In this work, variation of ingredients between couple of recipes are mined using closed itemsets extraction. The adaptation rules come from the interpretation of closed itemsets whose items correspond to the ingredients that have to be removed, kept, or added. This approach has been integrated as a wiki extension, providing a collaborative environment in which humans and machines may now collaborate to better acquire adaptation rules [38]. Humans (expert in cooking) may trigger automatic processes (knowledge discovery processes) and and may validate, using a specific user interface, proposition of adaptation rules as adaptation knowledge, which is then added to the knowledge base. In the same way, this environment integrates also the results of a new work on knowledge extraction where specific cooking adaptation rules (i.e. that can be applied to a single recipe) are generalized using close itemsets into generic adaptation rules, to make them usable on other recipes [60].

A module for acquiring a process semantic representation. While a process for acquiring cases from recipe preparation texts exists, the results are not perfect. In order for valid case representations to be available in the semantic wiki, a semi-automatic case acquisition tool was created [59]. This tools presents the user with a graphical interface through which it is able to interact with the case acquisition process. In order to limit the effort required, each correction entered by the user is propagated by the tool to the rest of the case representation.

Some other theoretical studies have been carried out that should be applied to some future versions of Taaable:

- The combination of workflows and interval algebras to represent procedural knowledge [55].
- The revision-based adaptation of cases represented in a qualitative algebra [41].
- The study of taxonomy merging [39]: several versions of the taxonomies used in Taaable (such as the food hierarchy) can be incoherent one with the others and a merging process is defined in order to obtain a consistent merged taxonomy.
- A continuous knowledge extraction process to ensure the non regression of the reasoning system according to the ontology evolution [50].

PAROLE Project-Team

6. New Results

6.1. Speech Analysis and Synthesis

Participants: Anne Bonneau, Vincent Colotte, Dominique Fohr, Yves Laprie, Joseph di Martino, Slim Ouni, Sébastien Demange, Fadoua Bahja, Agnès Piquard-Kipffer, Utpala Musti.

Signal processing, phonetics, health, perception,articulatory models, speech production, learning language, hearing help, speech analysis, acoustic cues, speech synthesis

6.1.1. Acoustic-to-articulatory inversion

6.1.1.1. Annotation of X-ray films and construction of articulatory models

Two databases have been annotated this year: one composed of 15 short sentences representing more than 1000 X-ray images and a second about CVCVs which has already been annotated by hand on sheets of papers. In the latter case we adapted tools of Xarticul software in order to enable a fast processing of these annotations.

Since images of the first database have been digitized from old films there are several spurious jumps and we thus developed tools to remove them during the construction of articulatory models. The big difference with previous databases processed is the presence of more consonants.

The articulatory model is supplemented by a clipping algorithm in order to take into account contacts between tongue and palate.

6.1.1.2. Articulatory copy synthesis

Acoustic features and articulatory gestures have always been studied separately. Articulatory synthesis could offer a nice solution to study both domains simultaneously. We thus explored how X-ray images could be used to synthesize speech. The first step consisted of connecting the 2D geometry given by mediosagittal images of the vocal tract with the acoustic simulation. Last year we thus developed an algorithm to compute the centerline of the vocal tract, i.e. a line which is approximately perpendicular to the wave front. The centerline is then used to segment the vocal tract into elementary tubes whose acoustic equivalents are fed into the acoustic simulation.

The frequency simulation enables the impact of local modifications of the vocal tract geometry to be evaluated easily. This is useful to investigate the contribution of the sagittal to area transformation in the synthetic speech spectrum. However, the sequence of area functions alone does not suffice to synthesize speech since consonants involve very fine temporal details (closure of the vocal tract and then release of the constriction for stops and fricatives for instance) which additionally have to be synchronized with the temporal evolution of the glottis area. Scenarii have thus been designed for VCV sequences and more generally for any consonant clusters. The idea consists of choosing relevant X-ray images near the VCV to be synthesized. These images can be duplicated just before the closure of the vocal tract, modified to simulate the constriction release for a stop...

This procedure has been applied successfully to copy sentences and VCV for four X-ray films of the DOCVACIM database http://www2i.misha.fr/flora/jsp/index.jsp. The next objective will be to develop a complete articulatory synthesis system.

6.1.1.3. Inversion from cepstral coefficients

The two main difficulties of inversion from cepstral coefficients are: (i) the comparison of cepstral vectors from natural speech and cepstral vectors generated by the articulatory synthesizer and (ii) the access to the articulatory codebook.

Last year we developed a bilinear frequency warping optimized to compensate for the articulatory model mismatch. However, the spectral tilt was not taken into account. We thus combined it with affine adaptation of the very first cepstral coefficients in order to take into account the spectral tilt. It turns out that the new adaptation enables a more relevant comparison of cepstral vectors since the geometric precision of the best solution is less than 1mm.

The second difficulty consists of exploring the articulatory codebook efficiently. Indeed, only a small number of hypercuboids could correspond to the input cepstral vector. The issue is to eliminate all cuboids, which cannot give rise to the input cepstral vector. This is easy when using formants as input data since all cuboids can be indexed easily with extreme values of formants. But this becomes impossible with cepstral vectors because the effect of the excitation source cannot be removed completely from cepstral coefficients. We thus use spectral peaks to access the codebook. However, there exist some spurious spectral peaks, and at the same time some peaks can be absent. We thus designed a lax matching between spectral peaks, which enables the comparison of a series of spectral peaks of the original speech with peaks calculated on synthetic speech. This matching algorithm allows the exploration to focus on 5% of the codebook instead of 40% when using only the peak corresponding to F2 is used.

6.1.1.4. Acoustic-to-articulatory inversion using a generative episodic memory

We have developed an episodic based inversion method. Episodic modeling is interesting for two reasons. First, it does not rely on any assumption about the mapping relationship between acoustic and articulatory, but rather it relies on real synchronized acoustic and articulatory data streams. Second, the memory structurally embeds the naturalness of the articulatory dynamics as speech segments (called episodes) instead of single observations as for the codebook based methods. Estimating the unknown articulatory trajectories from a particular acoustic signal, with an episodic memory, consists in finding the sequence of episodes, which acoustically best explains the input acoustic signal. We refer to such a memory as a concatenative memory (C-Mem) as the result is always expressed as a concatenation of episodes. Actually a C-Mem lacks from generalization capabilities as it contains only several examples of a given phoneme and fails to invert an acoustic signal, which is not similar to the ones it contains. However, if we look within each episode we can find local similarities between them. We proposed to take advantage of these local similarities to build a generative episodic memory (G-Mem) by creating inter-episodes transitions. The proposed G-Mem allows switching between episodes during the inversion according to their local similarities. Care is taken when building the G-Mem and specifically when defining the inter-episodes transitions in order to preserve the naturalness of the generated trajectories. Thus, contrary to a C-Mem the G-Mem is able to produce totally unseen trajectories according to the input acoustic signal and thus offers generalization capabilities. The method was implemented and evaluated on the MOCHA corpus, and on a corpus that we recorded using an AG500 articulograph. The results showed the effectiveness of the proposed G-Mem which significantly outperformed standard codebook and C-Mem based approaches. Moreover similar performances to those reported in the literature with recently proposed methods (mainly parametric) were reached.

The paradigm of episodic memories was also used for speech recognition. We do not extend the acoustic feature with any explicit articulatory measurements but instead we used the articulatory-acoustic generative episodic memories (G-mem). The proposed recognizer is made of different memories each specialized for a particular articulator. As all the articulators do not contribute equally to the realization of a particular phoneme, the specialized memories do not perform equally regarding each phoneme. We showed, through phone string recognition experiments that combining the recognition hypotheses resulting from the different articulatory specialized memories leads to significant recognition improvements.

6.1.2. Using Articulography for Speech production

Since we have an articulograph (AG500, Carstens Medizinelektronik) available, we can easily acquire articulatory data required to study speech production. The articulograph is used to record the movement of the tongue (this technique is called electromagnetography - EMA). The AG500 has a very good time resolution (200Hz), which allows capturing all articulatory dynamics. It has also a good precision. In fact, we performed recently an comparative study to assess the precision of the articulograph AG500 in comparison to a concurrent articulograph NDI Wave. In this study, we found that both systems presented similar results. We showed also that the accuracy is relatively independent of the sensor velocity, but decreases with the distance from magnetic center of the system [31].

To make the best use of the articulograph, we developed an original visualization software, VisArtico, which allows displaying the data acquired by an articulograph. It is possible to display the tongue contour and the lips contour animated simultaneously with acoustics. The software helps to find the midsagittal plane of the speaker and find the palate contour. In addition, VisArtico allows labeling phonetically the articulatory data[30].

We continuousely work on the usage this platform to acquire articulatory data that were used for articulatory-to-acoustic inversion but also to study the co-variation of speech clarity and coarticulatory patterns in Arabic [18]. The results revealed evident relationship between speech clarity and coarticulation: more coarticulation in formal speech and in strong prosodic position.

6.1.3. Speech synthesis

Visual data acquisition was performed simultaneously with acoustic data recording, using an improved version of a low-cost 3D facial data acquisition infrastructure. The system uses two fast monochrome cameras, a PC, and painted markers, and provides a sufficiently fast acquisition rate to enable an efficient temporal tracking of 3D points. The recorded corpus consisted of the 3D positions of 252 markers covering the whole face. The lower part of the face was covered by 70% of all the markers (178 markers), where 52 markers were covering only the lips so as to enable a fine lip modeling. The corpus was made of 319 medium-sized French sentences uttered by a native male speaker and corresponding to about 25 minutes of speech,.

We designed a first version of the text to acoustic-visual speech synthesis based on this corpus. The system uses bimodal diphones (an acoustic component and a visual one) and unit selection techniques (see 3.2.4). We have introduced visual features in the selection step of the TTS process. The result of the selection is the path in the lattice of candidates found in the Viterbi algorithm, which minimizes a weighted linear combination of three costs: the target cost, the acoustic joined cost, and the visual joined cost. Finding the best set of weights is a difficult problem by itself mainly because of their highly different nature (linguistic, acoustic, and visual considerations). To this end, we developed a method to determine automatically the weights applied to each cost, using a series of metrics that assess quantitatively the performance of synthesis.

The visual target cost includes visual and articulatory information. We implemented and evaluated two techniques: (1) Phonetic category modification, where the purpose was to change the current characteristics of some phonemes which were based on phonetic knowledge. The changes modified the target and candidate description for the target cost to better take into account their main characteristics as observed in the audio-visual corpus. The expectation was that their synthesized visual speech component would be more similar to the real visual speech after the changes. (2) Continuous visual target cost, where the visual target cost component is now considered as real value, and thus continuous, based on the articulatory feature statistics. This year, we continued working on improving the quality of the synthesis. This was done by continuously testing new strategies of weight tuning and improving our selection technique [26].

6.1.4. Phonemic discrimination evaluation in language acquisition and in dyslexia and dysphasia

6.1.4.1. Phonemic segmentation in reading and reading-related skills acquisition in dyslexic children and adolescents

Our computerized tool EVALEC was published [56] after the study of reading level and reading related skills of 400 hundred children from grade 1 to grade 4 (from age 6 to age 10) [58]. This research was supported by a grant from the French Ministery of Health (Contrat 17-02-001, 2002-2005). This first compurerized battery of tests in French language assessing reading and related skills (phonemic segmentation, phonological short term memory) comparing results both to chronological age controls and reading level age control in order to diagnostic Dyslexia. Both processing speed and accuracy scores are taken into account. This battery of tests is used by speech and langage therapists. We keep on examining the reliability (group study) and the prevalence (multiple case study) of 15 dyslexics' phonological deficits in reading and reading related skills in comparaison with a hundred reading level children [57], and by the mean of longitudinal studies of children from age 5 to age 17 [55]. This year, we started the development of a project which examined multimodal speech both with SLI, dyslexics and control children (30 children). Our goal is to examine visual contribution to speech perception accross differents experiments with a natural face (syllables with several conditions). Our goal is to search what can improve intelligibility in children who have sévère langague acquisition difficulties.

6.1.4.2. Langage acquisition and langage disabilities (deaf chidren, dysphasic children)

Providing help for improving French language acquisition for hard of hearing (HOH) children or for children with language disabilities was one of our goal : ADT (Action of Technological Development) Handicom [piquardkipffer:2010:inria-00545856:2]. The originality of this project was to combine psycholinguistical and speech analyses researchs. New ways to learn to speak/read were developed. A collection of three digital books has been written by Agnès Piquard-Kipffer for both 2-6, 5-9, 8-12 year old children (kindergarten, 1-4th grade) to train speaking and reading acquisition regarding their relationship with speech perception and audio-visual speech perception. A web interface has been created (using Symfony and AJAX technologies) in order to create others books for language impaired children. A workflow which transforms a text and an audio source in a video of digital head has been developed. This worklow includes an automatic speech alignment, a phonetic transcription, a speech synthetizer, a French cued speech coding and speaking digital head. A series of studies (simple cases studies, 5 deaf children and 5 SLI children and group studies with 2 kindergarten classes) were proposed to investigate the linguistical, audio-visual processing.... presumed to contribute to language acquisition in deaf children. Publication are submitted.

6.1.5. Enhancement of esophageal voice

6.1.5.1. Detection of F0 in real-time for audio: application to pathological voices

The work first rested on the CATE algorithm developed by Joseph Di Martino and Yves Laprie, in Nancy, 1999.The CATE (Circular Autocorrelation of the Temporal Excitation) algorithm is based on the computation of the autocorrelation of the temporal excitation signal which is extracted from the speech log-spectrum. We tested the performance of the parameters using the Bagshaw database, which is constituted of fifty sentences, pronounced by a male and a female speaker. The reference signal is recorded simultaneously with a microphone and a laryngograph in an acoustically isolated room. These data are used for the calculation of the contour of the pitch reference. When the new optimal parameters from the CATE algorithm were calculated, we carried out statistical tests with the C functions provided by Paul BAGSHAW. The results obtained were very satisfactory and a first publication relative to this work was accepted and presented at the ISIVC 2010 conference. At the same time, we improved the voiced / unvoiced decision by using a clever majority vote algorithm electing the actual F0 index candidate. A second publication describing this new result was published at the ISCIT 2010 conference. Recently we developed a new algorithm based on a wavalet transform applied to the cepstrum excitation. The resuts obtained were satisfactory. This work has been published in the ICMCS 2012 conference [14].

6.1.5.2. Voice conversion techniques applied to pathological voice repair

Voice conversion is a technique that modifies a source speaker's speech to be perceived as if a target speaker had spoken it. One of the most commonly used techniques is the conversion by GMM (Gaussian Mixture Model). This model, proposed by Stylianou, allows for efficient statistical modeling of the acoustic space of a speaker. Let "x" be a sequence of vectors characterizing a spectral sentence pronounced by the source speaker and "y" be a sequence of vectors describing the same sentence pronounced by the target speaker. The goal is to estimate a function F that can transform each source vector as nearest as possible of the corresponding target vector. In the literature, two methods using GMM models have been developed: In the first method (Stylianou), the GMM parameters are determined by minimizing a mean squared distance between the transformed vectors and target vectors. In the second method (Kain), source and target vectors are combined in a single vector "z". Then, the joint distribution parameters of source and target speakers is estimated using the EM optimization technique. Contrary to these two well known techniques, the transform function F, in our laboratory, is statistically computed directly from the data: no needs of EM or LSM techniques are necessary. On the other hand, F is refined by an iterative process. The consequence of this strategy is that the estimation of F is robust and is obtained in a reasonable lapse of time. This interesting result was published and presented at the ISIVC

2010 conference. Recently,we realized that one of the most important problems in speech conversion is the prediction of the excitation. In order to solve this problem we developed a new strategy based on the prediction of the ceptrum excitation pulses. This interesting result has been published in the SIIE 2012 conference [13].

6.1.5.3. Signal reconstruction from short-time Fourier transform magnitude spectra

Joseph Di Martino and Laurent Pierron developed in 2010 an algorithm for real-time signal reconstruction from short-time Fourier magnitude spectra. Such an algorithm has been designed in order to enable voice conversion techniques we are developing in Nancy for pathological voice repair. Recently Mouhcine Chami, an assistant-professor of the INPT institute at Rabat (Morocco) proposed a hardware implementation of this algorithm using FPGAs. This implementation has been publised in the SIIE 2012 conference [17].

6.1.6. Perception and production of prosodic contours in L1 and L2

6.1.6.1. Language learning (feedback on prosody)

A corpus, made up of 8 English sentences and 40 English isolated words has been recorded. Thirty three speakers pronounced the corpus under different conditions : without any audio feedback (first condition), with audio feedback (second condition, experiment realized one week after the first one). In order to test the permanence of the improvement due to feedback, a set of words and all the sentences were then pronounced without feedback (third condition, experiment realized after the second one). An English teacher helped us in the composition of the corpus and recorded it. Parts of this corpus have already been used to test the automatic speech alignment methods developed under the framework of ALLEGRO and implemented in jsnoori (ADT). The feedback will be progressively transferred from Winsnoori to Jsnoori.

6.1.6.2. Production of prosodic contour

The study of French contours (various types of continuations, end of sentences ...) confirmed the existence of patterns which are typical of French prosody. In order to determine the impact of French (the native language) on a second language pronunciation (English), a series of prosodic contours extracted from English sentences uttered by French speakers have been compared to French prosodic countours. To that purpose, French speakers recorded similar sentences in French and in English. Analysis of results is in progress. First results tend to show the impact of the native language ([15] and [10]).

6.1.7. Pitch detection

Over the last two years, we have proposed two new real time pitch detection algorithms (PDAs) based on the circular autocorrelation of the glottal excitation, weighted by temporal functions, derived from the CATE [53] original algorithm (Circular Autocorrelation of the Temporal Excitation), proposed initially by J. Di Martino and Y. Laprie. In fact, this latter algorithm is not constructively real time because it uses a post-processing technique for the Voiced/Unvoiced (V/UV) decision. The first algorithm we developed is the eCATE algorithm (enhanced CATE) that uses a simple V/UV decision less robust than the one proposed later in the eCATE+ algorithm.

We propose a recent modified version called the eCATE++ algorithm which focuses especially on the detection of the F0, the tracking of the pitch and the voicing decision in real time. The objective of the eCATE++ algorithm consists in providing low classification errors in order to obtain a perfect alignment with the pitch contours extracted from the Bagshaw database by using robust voicing decision methods. The main improvement obtained in this study concerns the voicing decision, and we show that we reach good results for the two corpora of the Bagshaw database. This algorithm is under a submission process in an international journal.

6.2. Automatic Speech Recognition

Participants: Sébastien Demange, Dominique Fohr, Christian Gillot, Jean-Paul Haton, Irina Illina, Denis Jouvet, Odile Mella, Luiza Orosanu, Othman Lachhab.

telecommunications, stochastic models, acoustic models, language models, automatic speech recognition, training, robustness

6.2.1. Core recognition

6.2.1.1. Broadcast News Transcription

A complete speech transcription system, named ANTS (see section 5.6), was initially developed in the framework of the Technolangue evaluation campaign ESTER for French broadcast news transcription. This year, in the context of the ETAPE evaluation campaign about transcription of radio and TV debates, the speech transcription system was improved. Large amounts of text data have been collected over the web. This new collected web data, plus new text and speech resources have made possible the creation and training of new acoustic models and new language models. Moreover new processing steps have been included in the transcription system, leading to much better performance than with the initial system. Several system variants have been developed, and for the ETAPE evaluation campaign, their results have been combined.

Extensions of the ANTS system have been studied, including the possibility to use the sphinx recognizers, and unsupervised adaptation processes. Training scripts for building acoustic models for the Sphinx recognizers are now available and take benefit of parallel computations on the computer cluster for a rapid optimization of the model parameters The Sphinx models are also used for speech/text alignment on both French and English speech data. A new speech transcription program has been developed for efficient decoding on the computer cluster, and easy modification of the decoding steps (speaker segmentation and clustering, data classification, speech decoding in one or several passes, ...). It handles both the Julius and Sphinx (versions 3 and 4) decoders.

This year, in the context of the ETAPE evaluation campaign, which deals with the transcription of radio and TV shows, mainly debates, the Julius-based and Sphinx-based transcription systems have been improved. Several system variants have been developed (relying on different features, and/or different normalization schemes, different processing steps, and different unsupervised adaptation processes); and, combining the output of the various systems led to significantly improved performance.

The recently proposed approach to grapheme-to-phoneme conversion based on a probabilistic method: Conditional Random Fields (CRF) was investigated further. CRF gives a long term prediction, and assumes a relaxed state independence condition. The proposed system was validated in a speech recognition context. Our approach compared favorably with the performance of the state-of-the-art Joint-Multigram Models (JMM) for the quality of the pronunciations, and it was also shown that combining the pronunciation variants generated by both the CRF-based and the JMM-based apporaches improves performance [21].

Concerning grapheme-to-phoneme conversion, a special attention was paid to infering the pronunciation variants of proper names [34], and the usage of additional information corresponding to the language origin of the proper name was investigated.

6.2.1.2. Non-native speakers

The performance of automatic speech recognition (ASR) systems drastically drops with non native speech. The main aim of non-native enhancement of ASRs is to make available systems tolerant to pronunciation variants by integrating some extra knowledge (dialects, accents or non-native variants).

Our approach is based on acoustic model transformation and pronunciation modeling for multiple non-native accents. For acoustic model transformation, two approaches are evaluated: MAP and model re-estimation. For pronunciation modeling, confusion rules (alternate pronunciations) are automatically extracted from a small non-native speech corpus. We presents [9] a novel approach to introduce confusion rules in the recognition system which are automatically learned through pronunciation modelling. The modified HMM of a foreign spoken language phoneme includes its canonical pronunciation along with all the alternate non-native pronunciations, so that spoken language phonemes pronounced correctly by a non-native speaker could be recognized. We evaluate our approaches on the European project HIWIRE non-native corpus which contains English sentences pronunced by French, Italian, Greek and Spanish speakers. Two cases are studied: the native language of the test speaker is either known or unknown. Our approach gives better recognition results than the classical acoustic adaptation of HMM when the foreign origin of the speaker is known. We obtain 22% WER reduction compared to the reference system.

6.2.1.3. Language Model

Christian Gillot has defended his Ph.D. thesis on the 17th September 2012. In his thesis, he proposes a new approach to estimate the language model probabilities for an automatic speech recognition system. The most commonly used language models in the state of the art are based on n-grams smoothed with Kneser-Ney method. Such models make use of occurrence counts of words sequences up to a maximum length (typically 5 words). These counts are computed on a huge training corpus. Christian's Ph.D. thesis starts by an empirical study of the errors of a state-of-the-art speech recognition system in French, which shows that there are many regular language phenomena that are out of reach of the n-gram models. This thesis thus explores a dual approach of the prevailing statistical paradigm by using memory models that process efficiently specific phenomena, in synergy with the n-gram models which efficiently capture the main trends in the corpus. The notion of similarity between long n-grams is studied in order to identify the relevant contexts to take into account in a first similarity language model. The data extracted from the corpus is combined via a Gaussian kernel to compute a new score. The integration of this non-probabilistic model improves the performance of a recognition system. A second model is then introduced, which is probabilistic and thus allows for a better integration of the similarity approach with the existing models. This second model improves the performance on texts in terms of perplexity. Some future works are further described, where the memory-based paradigm is transposed from the estimation of the n-gram probability up to the language model itself. The principle is to combine individual models together, where each model represents a specific syntactic structure, and also to combine these specific models with a standard n-gram model. The objective is to let specific models compensate for some weaknesses of n-gram models, which cannot capture sparse and rare phenomena, nor patterns that do not occur at all in the the training corpus. This approach hence opens new interesting perspectives in particular for domain adaptation.

6.2.1.4. Speech recognition for interaction in virtual worlds

Automatic speech recognition was investigated for vocal interaction in virtual worlds, in the context of serious games in the EMOSPEECH project. For training the language models, the text dialogs recorded by the TALARIS team (Midiki corpus) on the same serious game (but in a text-based interaction), have been manually corrected and used on addition of available broadcast news corpus. Different language models have then been created using different vocabulary sizes. The acoustic models were adapted from the radio broadcast news models, using state-of-the-art Maximum A Posteriori adaptation algorithm. This reduces the mismatch in recording conditions between the game devices and the original models trained on radio streams. A client-server speech recognition demonstrator has been developed. The client runs on an iPad; it records the speech input, sends it to the server, waits for the speech recognition answer, and finally displays the results. The server runs on a PC, relies on the sphinx4 decoder for decoding the received speech signal, and then sends the results to the iPad client.

6.2.2. Speech recognition modeling

Robustness of speech recognition to multiple sources of speech variability is one of the most difficult challenge that limits the development of speech recognition technologies. We are actively contributing to this area via the development of the following advanced modeling approaches.

6.2.2.1. Detailed modeling

Detailed acoustic modeling was further investigated using automatic classification of speaker data. With such an approach it is possible to go beyond the traditional four class models (male vs female, studio quality vs telephone quality). However, as the amount of training data for each class gets smaller when the number of classes increases, this limits the amount of classes that can efficiently be trained. Hence, we have investigated introducing a classification marging in the classification process. With such a marging, which handle boundary classification uncertainty, speech data at the class-boundary may belong to several classes. This increases the amount of training data in each class, which makes the class acoustic model parameters more reliable, and finally improved the overall recognition performance [22]. Combining maximum likelihood linear regression (MLLR) and maximum a posteriori (MAP) adaptation techniques leads to better speech recognition performance, and makes it possible to use more classes [35].

The approach was later improved by introducing a classification process which relies on phonetic acoustic models and the Kullback Leibler divergence measure to build maximally dissimilar clusters. This approach lead to better recognition results than the likelihood based classification approach used in previous experiments [20].

These class-based speech recognition systems were combined with more traditional gender-based system in the ETAPE campaign for the evaluation of speech transcription systems on French radio and TV shows.

6.2.2.2. Training HMM acouctic models

At the beginning of his second internship at Inria Nancy research laboratory, Othman Lachhab focused on the finalization of a speech recognition system based on context-independent HMMs models, using bigram probabilities for the phonotactic constraints and a model of duration following a normal distribution $\mathcal{N}(\mu, \sigma^2)$ incorporated directly in the Viterbi search process. Currently, he built a reference system for speaker-independent continuous phone recognition using Context- Independent Continuous Density HMM (CI-CDHMM) modeled by Gaussian Mixture Models (GMMs). In this system he developed his own training technique, based on a statistical algorithm estimating the classical optimal parameters. This new training process compares favorably with already published HMM technology on the same test corpus (TIMIT) and has been published in the ICMCS 2012 conference [23].

6.2.3. Speech/text alignment

6.2.3.1. Evaluation of speech/text alignment tools

Speech-text alignment tools are frequently used in speech technology and research: for instance, for training or assessing of speech recognition systems, the extraction of speech units in speech synthesis or in foreign language learning. We designed the software CoALT (Comparing Automatic Labelling Tools) for comparing two automatic labellers or two speech-text alignment tools, ranking them, and displaying statistics about their differences.

The main feature of CoALT is that a user can define its own criteria for evaluating and comparing the speech-text alignment tools since the required quality for labelling depends on the targeted application. Beyond ranking, our tool provides useful statistics for each labeller and above all about their differences and can emphasize the drawbacks and advantages of each labeller. We have applied our software for the French and English languages [19] but it can be used for another language by simply defining the list of the phonetic symbols and optionally a set of phonetic rules.

6.2.3.2. Alignment with non-native speech

Non-native speech alignment with text is one critical step in computer assisted foreign language learning. The alignement is necessary to analyze the learner's utterance, in view of providing some prosody feedback (as for example bad duration of some syllables - too short or too long -). However, non-native speech alignement with text is much more complicated than native speech alignment. This is due to the pronunciation deviations observed on non-native speech, as for example the replacement of some target language phonemes by phonemes of the mother tongue, as well as errors in the pronunciations. Moreover, these pronunciation deviations are strongly speaker dependent (i.e. they depend on the mother tongue of the speaker, and on its fluency in the target foreign lanaguage) which makes their prediction difficult.

However, the first step in automatic computer assisted language learning is to check that the pronunced word or utterance corresponds to the expected sentence, otherwise, if the user has not pronunced the correct words it is useless to proceed further with a detailed analysis of the pronunciation to check for possible misspronunciations. In order to decide if the pronunced utterance corresponds to the expected word or sentence, a force phonetic alignment of the sentence is compared to free decoding of the same sentence. Several comparison features are then defined, such as the number of matching phonemes, the percentage of frames having the save category label, ..., as well as the likelihood ratio. A classifier is then used to decide whether text and speech utterance match or not [36], [28].

These non-native phonetic alignments processes developed in the framework of the ALLEGRO project are currently under implementation in the JSNOORI software, and the processing should be completed by the developpement of automatic feedback procedures.

6.3. Speech-to-Speech Translation and Langage Modeling

Participants: Kamel Smaïli, David Langlois, Sylvain Raybaud, Motaz Saad, Denis Jouvet, Cyrine Nasri.

machine translation, statistical models

Sylvain Raybaud has just defended his thesis untitled “De l’utilisation de mesures de confiance en traduction automatique : évaluation, post-édition et application à la traduction de la parole.”. His contributions are the following: study and evaluation of confidence measures for Machine Translation, an original algorithm to automatically build an artificial corpus with errors for training the confidence measures, development of an entire speech-to-text translation system.

In the scope of Confidence Measures, we participated to the World Machine Translation evaluation campaign (WMT2012 http://www.statmt.org/wmt12/quality-estimation-task.html). More precisely, we proposed a Quality Estimation system to the Quality Estimation shared task. The goal was to predict the quality of translations generated by an automatic system. Each translated sentence is given a score between 1 and 5. The score is obtained using several numerical or boolean features calculated according to the source and target sentences. We perform a linear regression of the feature space against scores in the range [1:5]. To this end, we use a Support Vector Machine. We experiment with two kernels: linear and radial basis function. In our system we use the features from the shared task baseline system and our own features (based on the work from the Sylvain Raybaud’s thesis). This leads to 66 features. To deal with this large number of features, we propose an in-house feature selection algorithm. Our system came 5th among 19 systems. This work was publish in [24]. In the continuation of this research, we contributed to the development of a Quality Estimation tool (quest: https://github.com/lspecia/quest). For that, David Langlois was invited by Lucia Specia at University of Sheffield, Computer Sciences department, Natural Language Processing group. We added our own features into quest. This tool is dedicated to be available for the research community.

Another objective of our research work, with the Cyrine Nasri’s Phd thesis, is to retrieve bilingual phrases for machine translation. As in fact, current statistical machine translation systems usually build an initial word-to-word alignment before learning phrase translation pairs. This operation needs many matching between different single words of both considered languages. We propose a new approach for phrase-based machine translation which does not need any word alignments. It is based on inter-lingual triggers determined by Multivariate Mutual Information. This algorithm segments sentences into phrases and finds their alignments simultaneously. Inspite of the youth of this method, experiments showed that the results are competitive but needs some more efforts in order to overcome the one of state-of-the-art methods.

Another aspect of the research of the group is to work on under resourced language related to Arabic. In fact, in several countries through the Arabic world, only few people speak the modern standard Arabic language. People speak something which is inspired from Arabic but could be very different from the modern standard Arabic. This one is reserved for the official broadcast news, official discourses and so on. The study of dialect is more difficult than any other natural language because it should be noted that this language is not written. A preliminary work has been done knowing that our final objective is to propose a machine translation between the different Arabic dialects and modern standrad Arabic. This issue is very difficult and challenging because no corpus does exist, vernaculars are different even within the same country, etc.

Last, Motaz Saad has started his thesis in November 2011. His objective is to work on opinion analysis in multilingual documents from internet. During this year, he retrieved comparable corpus from the web, and proposed a method to align these corpora at document level. He proposed algorithms to measure the degree of comparability between documents. He submitted his work to the International Conference on Corpus Linguistics (CICL2013).

In the framework of the ETAPE evaluation campaign a new machine learning based process was developed to select the most relevant lexicon to be used for the transcription of the speech data (radio and TV shows). The approach relies on a neural network trained to distinguish between words that are relevant for the task and those that are not. After training, the neural network (NN) is applied to each possible word (extracted from a very large text corpus). Then the words that have the largest NN output score are selected for creating the speech recognition lexicon. Such an approach can handle counts of occurences of the words in various data subsets, as well as other complementary informations, and thus offer more perspectives than the traditional unigram-based selection procedures.

PERCEPTION Team

6. New Results

6.1. 3D shape analysis and registration

We address the problem of 3D shape registration and we propose a novel technique based on spectral graph theory and probabilistic matching. Recent advancement in shape acquisition technology has led to the capture of large amounts of 3D data. Existing real-time multi-camera 3D acquisition methods provide a frame-wise reliable visual-hull or mesh representations for real 3D animation sequences The task of 3D shape analysis involves tracking, recognition, registration, etc. Analyzing 3D data in a single framework is still a challenging task considering the large variability of the data gathered with different acquisition devices. 3D shape registration is one such challenging shape analysis task. The main contribution of this chapter is to extend the spectral graph matching methods to very large graphs by combining spectral graph matching with Laplacian embedding. Since the embedded representation of a graph is obtained by dimensionality reduction we claim that the existing spectral-based methods are not easily applicable. We discuss solutions for the exact and inexact graph isomorphism problems and recall the main spectral properties of the combinatorial graph Laplacian; We provide a novel analysis of the commute-time embedding that allows us to interpret the latter in terms of the PCA of a graph, and to select the appropriate dimension of the associated embedded metric space; We derive a unit hyper-sphere normalization for the commute-time embedding that allows us to register two shapes with different samplings; We propose a novel method to find the eigenvalue-eigenvector ordering and the eigenvector sign using the eigensignature (histogram) which is invariant to the isometric shape deformations and fits well in the spectral graph matching framework, and we present a probabilistic shape matching formulation using an expectation maximization point registration algorithm which alternates between aligning the eigenbases and finding a vertex-to-vertex assignment. See [22], [34], [19] for more details.

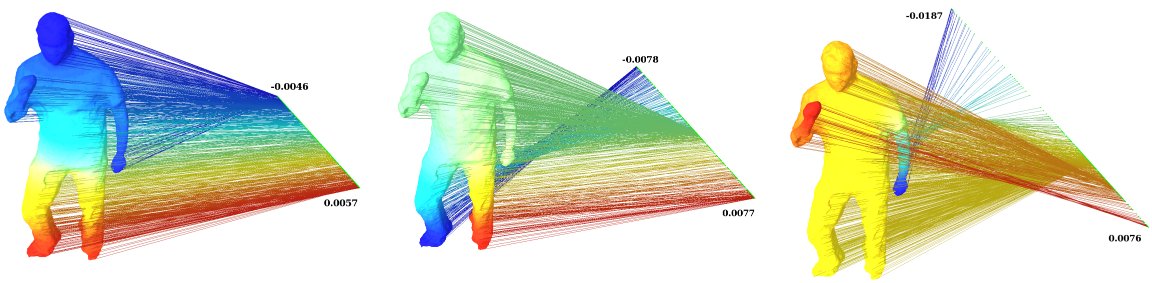


Figure 5. This is an illustration of the concept of the PCA of a shape embedding. The shapes's vertices are projected onto the second, third and fourth eigenvectors of the Laplacian matrix. These eigenvectors can be viewed as the principal directions of the shape (see [34] for more details).

6.2. High-resolution depth maps based on TOF-stereo fusion

The combination of range sensors with color cameras can be very useful for a wide range of applications, e.g., robot navigation, semantic perception, manipulation, and telepresence. Several methods of combining range- and color-data have been investigated and successfully used in various robotic applications. Most of these systems suffer from the problems of noise in the range-data and resolution mismatch between the range sensor and the color cameras, since the resolution of current range sensors is much less than the resolution of color cameras. High-resolution depth maps can be obtained using stereo matching, but this often fails to construct accurate depth maps of weakly/repetitively textured scenes, or if the scene exhibits complex self-occlusions. Range sensors provide coarse depth information regardless of presence/absence of texture. The use of a calibrated system, composed of a time-of-flight (TOF) camera and of a stereoscopic camera pair, allows data fusion thus overcoming the weaknesses of both individual sensors. We propose a novel TOF-stereo fusion method based on an efficient seed-growing algorithm which uses the TOF data projected onto the stereo image pair as an initial set of correspondences. These initial “seeds” are then propagated based on a Bayesian model which combines an image similarity score with rough depth priors computed from the low-resolution range data. The overall result is a dense and accurate depth map at the resolution of the color cameras at hand. We show that the proposed algorithm outperforms 2D image-based stereo algorithms and that the results are of higher resolution than off-the-shelf color-range sensors, e.g., Kinect. Moreover, the algorithm potentially exhibits real-time performance on a single CPU. See [27], [33] for more details.

6.3. Simultaneous sound-source separation and localization

Human-robot communication is often faced with the difficult problem of interpreting ambiguous auditory data. For example, the acoustic signals perceived by a humanoid with its on-board microphones contain a mix of sounds such as speech, music, electronic devices, all in the presence of attenuation and reverberations. We proposed a novel method, based on a generative probabilistic model and on active binaural hearing, allowing a robot to robustly perform sound-source separation and localization. We show how interaural spectral cues can be used within a constrained mixture model specifically designed to capture the richness of the data gathered with two microphones mounted onto a human-like artificial head. We describe in detail a novel expectation-maximization (EM) algorithm that alternates between separation and localization, we analyse its initialization, speed of convergence and complexity, and we assess its performance with both simulated and real data. Subsequently, we studied the *binaural manifold*, i.e., the low-dimensional space of sound-source locations embedded in the high-dimensional space of perceived interaural spectral features, and we provided a method for mapping interaural cues onto source locations. See [25], [24], [26]

6.4. Sound localization and recognition with a humanoid robot

We addressed the problem of localizing recognizing everyday sound events in indoor environments with a consumer robot. For localization, we use the four microphones that are embedded into the robot’s head. We developed a novel method that uses four non-coplanar microphones and that guarantees that for each set of pairwise TDOA (time difference of arrival) there is a unique 3D source location. For recognition, sounds are represented in the spectrotemporal domain using the stabilized auditory image (SAI) representation. The SAI is well suited for representing pulse-resonance sounds and has the interesting property of mapping a time-varying signal into a fixed-dimension feature vector space. This allows us to map the sound recognition problem into a supervised classification problem and to adopt a variety of classifications schemes. We developed a complete system that takes as input a continuous signal, splits it into significant isolated sounds and noise, and classifies the isolated sounds using a catalogue of learned sound-event classes. The method is validated with a large set of audio data recorded with a humanoid robot in a typical home environment. Extended experiments showed that the proposed method achieves state-of-the-art recognition scores with a twelve-class problem, while requiring extremely limited memory space and moderate computing power. A first real-time embedded implementation in a consumer robot show its ability to work in real conditions. See [23], [28] for more details.

6.5. Audiovisual fusion based on a mixture model

The problem of multimodal clustering arises whenever the data are gathered with several physically different sensors. Observations from different modalities are not necessarily aligned in the sense there there is no obvious way to associate or to compare them in some common space. A solution may consist in considering multiple clustering tasks independently for each modality. The main difficulty with such an approach is to guarantee that the unimodal clusterings are mutually consistent. In this paper we show that multimodal clustering can be addressed within a novel framework, namely conjugate mixture models. These models exploit the explicit transformations that are often available between an unobserved parameter space (objects) and each one of the observation spaces (sensors). We formulate the problem as a likelihood maximization task and we derive the associated expectation-maximization algorithm. The algorithm and its variants are tested and evaluated within the task of 3D localization of several speakers using both auditory and visual data. See [36], [30], [29] for more details.

POTIOC Team

6. New Results

6.1. Spatial augmented reality for physical drawing

Participants: Jérémy Laviole, Martin Hachet.

Spatial augmented reality (SAR) promises the integration of digital information in the real (physical) world through projection. We proposed different tools to improve speed or ease the drawing by projecting photos, virtual construction lines and interactive 3D scenes (published in the 3DUI symposium [16]). We explored the creation of tools which help to create drawings that are "difficult" to achieve for a human being, such as stereographic drawings (published in the 3DCHI CHI workshop [18]). Through these tools, we want to apply existing computer graphics techniques to enhance existing drawing tools, and to use it to teach how to draw. Furthermore, we proposed some insights for the creation of digital games and programs which can take full advantages of physical drawings (published in the UIST doctoral symposium [17]).

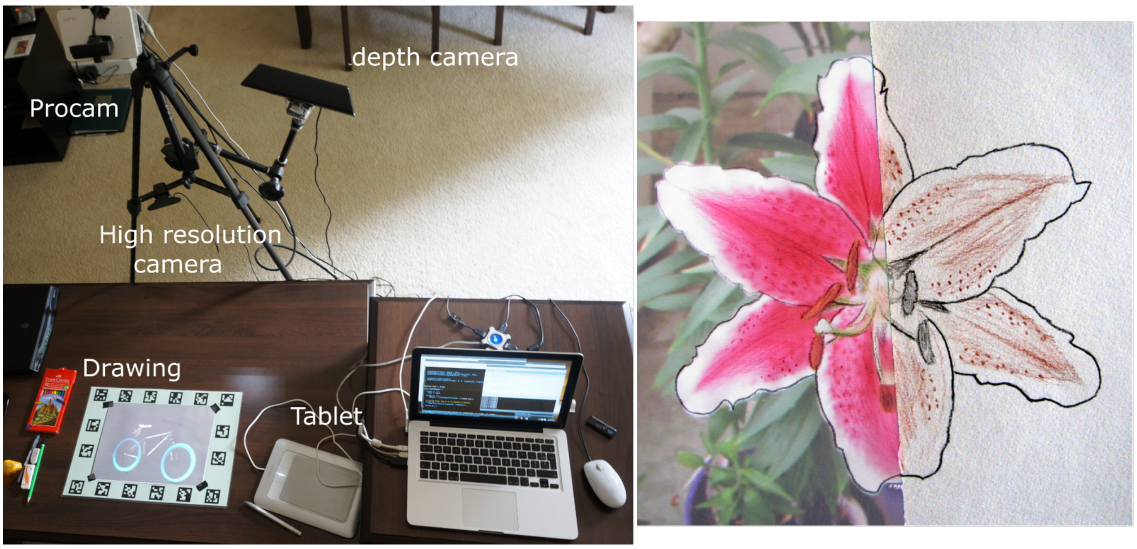


Figure 3. Left: Spatial augmented reality system for physical drawing. Right: Projection of a source image on overlay of an actual drawing, to teach drawing thanks to computer graphic tools [16].

6.2. Brain-Computer Interfaces

Participants: Fabien Lotte, Florian Larrue, Martin Hachet.

As part of our research on Brain-Computer Interfaces (BCI), our contributions addressed two different levels: 1) the brain signal processing level, in order to design more efficient BCI systems and 2) the applications level, in order to propose and explore new BCI applications.

At the signal processing level, we explored and designed new features to represent ElectroEncephaloGraphic (EEG) signals. In particular we explored multifractal cumulants and predictive complexity features (which we published in the Neurocomputing journal [5]), as well as waveform length features together with an optimal spatial filter that we designed for such features (which we published in the ICPR international conference [19]). All these features proved useful to classify EEG signals, and, more importantly, increased the classification performances of the system when combined together with the gold standard features, namely, band power features. Thus, this contributed to extending the repertoire of features available to BCI designers as well as increasing BCI performances. Nevertheless, our studies of BCI and educational research led us to the conclusion that current BCI feedback training approaches (which aimed at teaching people how to use a BCI and how to control their own brain activity), are most probably highly inappropriate and one of the major causes for the limited performances of current BCI - maybe more than signal processing methods. We therefore stressed the need for alternative feedback training approaches for BCI in a publication at the international BBCI workshop [20].

At the application level, we mostly focused on Virtual Reality (VR) related applications. Indeed, together with other groups in the field, we reviewed how BCI and VR could be combined in order to give rise to new applications and to improve BCI designs. This was published in a book chapter dedicated to BCI [22]. Similarly, with international colleagues, we reviewed and envisioned new applications of BCI outside the medical domain, and proposed guidelines to move towards these new applications. This notably includes VR and game applications, user-state monitoring, neuro-evaluation, training and education, cognitive improvement as well as safety and security. This was published in the IEEE Computer journal [8]. Finally, we proposed a new and innovative application of BCI: using it as a tool to study spatial cognition and transfer from VR to real environments. In particular, since BCI can be used to navigate a Virtual Environment (VE) without any motor activity, BCI can be used to assess how much motor activity is really needed to transfer spatial knowledge from a VE to a real one. This is what we did by comparing a BCI and a treadmill in order to teach users a path in a VE and then asking them to retrieve this path in the real world. Contrary to what was believed before, our results showed that motor activity is not necessary to learn a path in VR. We showed that what is really necessary is performing an action, but that this action does not have to be motor, and can be, for instance, cognitive (e.g., imagining hand movements), with a BCI. This was published in the VRST international conference [14].

6.3. Understanding user gestures for touch screen-based 3D User Interfaces

Participants: Aurélie Cohé, Martin Hachet.

In the scope of the ANR project Instinct, we studied how users tend to interact with a touchscreen for interacting with 3D content. Our main contributions were to study user behaviors with a standard touchscreen on the one hand, and with a pressure sensitive touchscreen on the other hand.

Multi-touch interfaces have emerged with the widespread use of smartphones. Although a lot of people interact with 2D applications through touchscreens, interaction with 3D applications remains little explored. Most 3D object manipulation techniques have been created by designers who have generally put users aside from the design creation process. We conducted a user study to better understand how non-technical users tend to interact with a 3D object from touchscreen inputs. The experiment has been conducted while users were manipulating a 3D cube with three viewpoints for rotations, scaling and translations (RST). Sixteen users participated and 432 gestures were analyzed. To classify data, we introduce a taxonomy for 3D manipulation gestures with touchscreens. Then, we identify a set of strategies employed by users to perform the proposed cube transformations. Our findings suggest that each participant uses several strategies with a predominant one. Furthermore, we conducted a study to compare touchscreen and mouse interaction for 3D object manipulations. The results suggest that gestures are different according to the device, and touchscreens are preferred for the proposed tasks. Finally, we propose some guidelines to help designers in the creation of more user friendly tools. This work was published in the Graphics Interface (GI) conference [12] as well as in the Computers and Graphics journal [6].

Figure 4. A user navigating a virtual model of the city of Bordeaux with a BCI, in order to learn a specific path [14].

Moreover, few works have focused on the relation between the manipulated data and the quantity of force applied with the fingers sliding on a touch sensor. In another work, we conducted two user studies to better understand how users manage to control pressure, and how they tend to use this input modality. A first set of experiments allows us to characterize pressure in relation to finger motions. Based on the results of this study, we designed a second set of experiments focusing on the completion of 3D manipulation tasks from 2D gestures. The results indicate that a strong relationship exists between the actions the participants intend to perform, and the quantity of force they apply for 3D object manipulations. This finding opens new promising perspectives to enhance user interfaces dedicated to force-based touch sensors.

All these works were published in the PhD thesis of Aurélie Cohé [4], which was defended on December 13th, 2012.

6.4. Virtual reality for Musical Performance

Participants: Florent Berthaut, Martin Hachet.

Immersive virtual environments open new perspectives for music interaction, notably for the visualization of sound processes and of musical structures, for the navigation in musical compositions, for the manipulation of sound parameters and for musical collaboration. Research conducted by Florent Berthaut and Martin Hachet, in collaboration with Myriam Desainte-Catherine from the SCRIME/LaBRI, explore these new possibilities.

Among the current projects, development of the Drile immersive virtual musical instrument was pursued in order to enable various scenographic setups that will be evaluated in the context of public performance. New perspectives for the Tunnels, 3D widgets for musical modulation (see Figure 6), were published as a Poster in the Proceedings of the Symposium on 3D User Interfaces (3DUI) [10]. Novel 3D selection techniques that take music interaction constraints into account are also being designed.

Another project was conducted with David Janin and Benjamin Martin from the LaBRI on new musical models that will be used to improve the hierarchical musical structures manipulated with Drile. It was published in the International Conference on Semantic Computing [11].

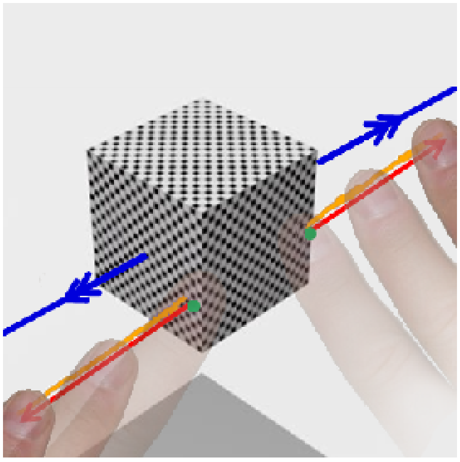


Figure 5. Analysis of users' gestures on touch screen to manipulate 3D content [6].

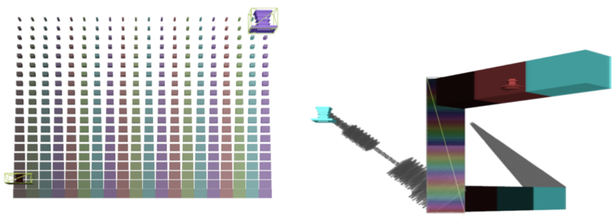


Figure 6. The Tunnels 3D widgets for musical modulation.

A collaboration was started with researchers of the Center for Computer Research on Music and Acoustics (CCRMA) of Stanford University. Florent Berthaut was invited for two months at CCRMA, where he worked with Luke Dahl and Chris Chafe on the implementation of musical collaboration modes in immersive virtual environments. A first result is the design of 3D musical collaboration widgets for Drile, which will be evaluated with musicians.

Another project was initiated with researchers of the Bristol Interaction and Graphics group of the University of Bristol. This project aims at improving the audience experience with Digital Musical Instruments (DMIs). These instruments are often confusing for spectators because of the variety of used components and because of the lack of physical continuity between musicians gestures and the resulting sound. A novel approach was implemented using a mixed-reality system in order to reveal the mechanisms of DMIs (see Figure 7). A description of this approach and of the first prototype will be submitted to the conference on New Interfaces for Musical Expression.

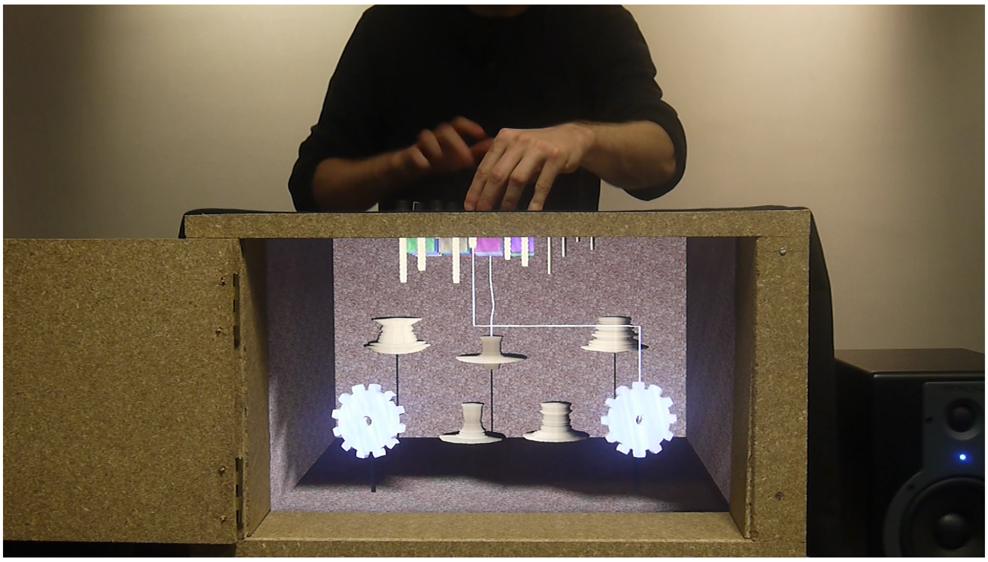


Figure 7. Rouages: a mixed-reality system that reveals the mechanisms of digital musical instruments to the audience.

6.5. Gateway driving simulator

Participants: Florian Larrue, Pauline Davignon, Pierre-Alexandre Favier, Martin Hachet.

As part of the SIMCA FUI project, the POTIOC team focuses on the design and evaluation of a gateway driving simulator, to teach drivers how to drive an airport gateway in virtual reality, i.e., in a safe and cost-effective environment. Gateways are the means to transfer passengers between the airport and the plane, for departures and arrivals. We have developped 3 simulators with different immersion levels (small, medium and immersive simulators, see, e.g., Figure 8). For each immersion level, we developped protocols in order to evaluate the impact of 3D technologies such as stereoscopy and head tracking on users' performances and preferences. Experimentations and evaluations are currently in progress.

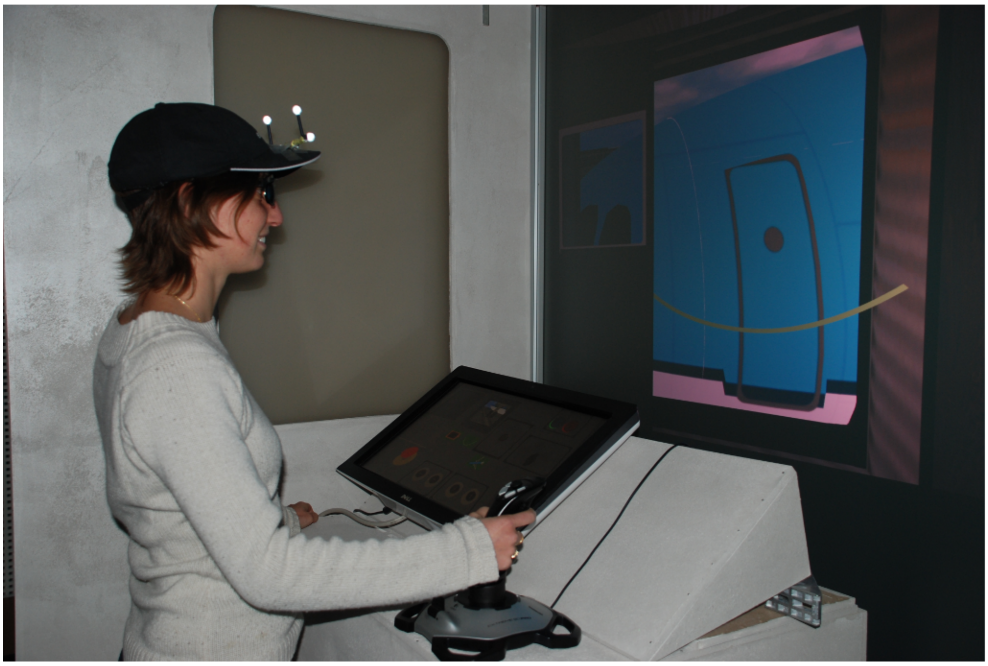


Figure 8. A user, equipped with head tracking and stereoscopic glasses, using the gateway driving simulator.

PRIMA Project-Team

6. New Results

6.1. Social behaviors recognition

Participants: Wafa Benkaouar, Claudine Combe, Dominique Vaufreydaz [correspondant].

Figure 5. On the left image, one can see the telemeter range in red, the foot detection (blue spot) and the angle view from the Kinect (in green). the middle and right image represent RGB camera en depth view from the Kinect.

Recognition of social behaviors is an unconscious innate cognitive process vital to human communication. This skill enables anticipation and increases interactive exchanges quality between humans. Among social behaviors, engagement is the expression of intention for interaction. During engagement phase, many non-verbal signals are used to communicate this intention to the partner, e.g. posture, gaze, spatial information, gestures, vocal cues. Within the context of frail or elderly people at home, companion robots must also be able to detect the engagement of humans in order to adapt their responses during interaction with humans to increase their acceptability.

Classical approaches in the domain are dealing with spatial information. Our hypothesis was that relative spatial information of people and robot are not discriminative in a home-like environment [15]. Our approach integrates multimodal features gathered using a robot companion equipped with a Kinect from Microsoft (see figure 5). Confronted to a robot centered dataset for multimodal social signal processing recorded in a home-like environment, the evaluation highlights its robustness and validates use of such technique in real environment (50% of error reduction). Our experimentations also confirm results from cognitive science domain [61].

6.2. Live monitoring and correction of 3DTV broadcasts

Participants: Pierre Arquier, Frédéric Devernay [correspondant], Sylvain Duchêne, Sergi Pujades-Rocamora, Matthieu Volat.

6.2.1. 3D broadcast monitoring and correction:

One of the achievements of the 3DLive FUI project was the transfer of real-time 3D video monitoring and correction algorithms to the Binocle company, and their integration into the TaggerLive product, which was used during several 3DTV broadcasts between 2010 and 2012 for live monitoring and correction of stereoscopic video. The algorithms that were developed within the PRIMA team and transferred into the TaggerLive are:

Multiscale view-invariant feature detection and matching on the GPU.

Computation of a temporally smooth and robust correction (or rectification) to remove the vertical disparity in the stereoscopic video while keeping the image aspect.

Real-time monitoring of the “depth budget”, or the histogram of the horizontal disparity;

Live alerts when stereoscopic production rules are broken, such as when the disparities are too large, or when there is a stereoscopic window violation.

Real-time implementation of a state-of-the-art dense stereo matching method on the GPU.

6.2.2. 3D content adaptation:

3D shape perception in a stereoscopic movie depends on several depth cues, including stereopsis. For a given content, the depth perceived from stereopsis highly depends on the camera setup as well as on the display size and distance. This can lead to disturbing depth distortions such as the cardboard effect or the puppet theater effect. As more and more stereoscopic 3D content is produced in 3D (feature movies, documentaries, sports broadcasts), a key point is to get the same 3D experience on any display. For this purpose, perceived depth distortions can be resolved by performing view synthesis. We have proposed [19] a real time implementation of a stereoscopic player based on the open-source software Bino, which is able to adapt a stereoscopic movie to any display, based on user-provided camera and display parameters.

6.2.3. Focus mismatch detection:

Live-action stereoscopic content production requires a stereo rig with two cameras precisely matched and aligned. While most deviations from this perfect setup can be corrected either live or in post-production, a difference in the focus distance or focus range between the two cameras will lead to unrecoverable degradations of the stereoscopic footage. We have developed a method [18] to detect focus mismatch between views of a stereoscopic pair in four steps. First, we compute a dense disparity map. Then, we use a measure to compare focus in both images. After this, we use robust statistics to find which images zones have a different focus. Finally, to give useful feedback, we show the results on the original images and give hints on how to solve the focus mismatch.

6.3. Simultaneous localization and mapping (SLAM)

Participants: James Crowley, Marion Decrouez, Frédéric Devernay.

Localisation, place recognition, object recognition. Live processing of a video sequence taken from a single camera enables to model an a priori unknown 3D scene. Metrical SLAM (Simultaneous Localization and Mapping) algorithms track the camera pose while reconstructing a sparse map of the visual features of the 3D environment. Such approaches provide the geometrical foundation for many augmented reality applications in which informations and virtual objects are superimposed on live images captured by a camera. Improving such systems will enable in the future precise industrial applications such as guided-maintenance or guided-assembly in wide installations.

A problem with current methods is the assumption that the environment is static. Indoor environments such as supermarket ailes and factory floors may contain numerous objects that are likely to be moved, disrupting a localization and mapping system. We explore methods for automatic detection and modeling of such objects. We define the scene as a static structure that may contain moving objects and objects are defined as a set of visual features that share a common motion compared to the static structure [39]. Using several explorations of a camera in the same scene, we detect and model moved objects while reconstructing the environment. Experiments highlignt the performance of the method in a real case of localization in an unknown indoor environment.

6.4. Post-production tools for 3-D Cinema

Participants: Laurent Boiron, Frédéric Devernay [correspondant], Sylvain Duchêne, Sergi Pujades-Rocamora.

Over the past 6 years, we have been developing 3D movie processing techniques which have been used for the production and post-production of 3D movies (mainly feature-length movies, documentaries and commercials). These include image alignment, view interpolation, depth map computation, etc. These algorithms were developed as C++ libraries, and can be executed using standalone tools. Since the movie post-production workflow relies mainly on standard tools for compositing, color grading, etc., and these tools can be extended by plugin mechanisms, we integrated our post-production algorithms into such a tool, namely Nuke by The Foundry.

We also developed a new method for stereoscopic video cut and paste. Video cut-and-paste consists in semi-interactively segmenting a video object from a video stream, and pasting the segmented video object in another video. The object segmentation is done using a small number of strokes made on a few frames of the video, and can be corrected interactively. Existing methods only worked on monoscopic videos, and extending it to stereoscopic videos required solving important challenges:

- The video object must not only remain consistent over time, but also between the left and right views.
- The video object may be partially occluded in one or both views.
- The camera setup may be different between the first and the second video, causing depth distortion or different depth effects.

We solved the first two challenges by adding left-right stereo consistency based on dense stereo matching, as well as temporal consistency based on optical flow, in an optimization framework based on graph cuts. The user interface was also taken into consideration in the algorithm: any correction of the results (i.e. new strokes on an image) will only propagate forward in time.

6.5. Scene flow computation from RGBZ data

Participants: Frédéric Devernay [correspondant], Julian Quiroga.

The scene flow describes the motion of each 3D point between two times steps. With the arrival of new depth sensors, as the Microsoft Kinect, it is now possible to compute scene flow with a single camera, with promising repercussion in a wide range of computer vision scenarios. We proposed [22] a novel method to compute scene flow by tracking in a Lucas-Kanade framework. Scene flow is estimated using a pair of aligned intensity and depth images, but rather than computing a dense scene flow as in most previous methods, we get a set of 3D motion fields by tracking surface patches. Assuming a 3D local rigidity of the scene, we propose a rigid translation flow model that allows to solve directly for the scene flow by constraining the 3D motion field both in intensity and depth data. In our experimentation we achieve very encouraging results. Since this approach solves simultaneously for the 2D tracking and for the scene flow, it can be used for action recognition in existing 2D tracking based methods or to define scene flow descriptors.

6.6. Knit your Ideas Into Smart Spaces

Participants: Joelle Coutaz, Alexandre Demeure [correspondant], Emeric Fontaine.

We developped KISS (Knit your Ideas Into Smart Spaces), an end-user development system for the home. KISS enables users to program their home with sentences expressed in a pseudo-natural language. Programs can be tested either with the virtual home or in the real home. We led an evaluation that shows that users are able to program a real-life scenario. This work is described in the phd manuscript of Emeric Fontaine [46]. An experimental evaluation shows that KISS can be used to programm a real life scenario.

Particpants encountered some difficulties related to the restricted vocabulary used for the experiment. Some difficulties also occured relative to the understanding of "progressive verbs". To overcome these problems, we envision a system for co-constructing vocabulary with the system, which may lead to the definition of multiple language for communicating with the system.

6.7. Attention-Based Navigation

Participants: Thomas Fisher, Thierry Fraichard [correspondant], Patrick Reignier.

Assistant robots and robot companions are designed to share the human living space, to navigate among and interact with human beings. From the mobility point of view, roboticists have recently striven to develop navigation scheme geared towards achieving so-called "socially acceptable motions". To that end, various concepts borrowed from environmental psychology and anthropology have been used, the "personal space" concept from Proxemics being perhaps the most widely used.

The purpose of our work here is to further the research in this area by taking into account other factors such as human activities, interaction configurations and intentions. An attentional model derived from cognitive psychology is used to dynamically determine the "focus of attention" of the persons involved in a given task. Depending on the task at hand, the robot uses the attention information in order to decide its future course of action so as, for instance, to attract one person's attention or, on the contrary, to minimize the disturbance caused.

REVES Project-Team

5. New Results

5.1. Plausible Image Rendering

5.1.1. Rich Intrinsic Image Decomposition of Outdoor Scenes from Multiple Views

Participants: Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

Intrinsic image decomposition aims at separating photographs into independent reflectance and illumination layers. We show that this ill-posed problem can be solved by using multiple views of the scene from which we derive additional constraints on the decomposition.

Our first method uses pictures from multiple views at a *single time of day* to automatically reconstruct a 3D point cloud of an outdoor scene. Although this point cloud is sparse and incomplete, it is sufficient to compute plausible sky and indirect illumination at each oriented 3D point, given an environment map that represents incoming distant radiance. We introduce an optimization method to estimate sun visibility over the point cloud, which compensates for the lack of accurate geometry and allows the extraction of precise cast shadows. We finally use image-guided propagation algorithms to propagate the illumination computed over the sparse point cloud to every pixel, and to separate the illumination into distinct sun, sky, and indirect components. This *rich intrinsic image decomposition* enables advanced image manipulations, illustrated in Figure 3 .

This work has led to the RID software (Section 4.1) and to a technology transfer agreement with Autodesk (Section 6.1.1.1). A paper will be published in the IEEE Transactions on Visualization and Computer Graphics journal [18] (in press). It has also been presented at SIGGRAPH 2012 in the Poster and Talk sessions [22].

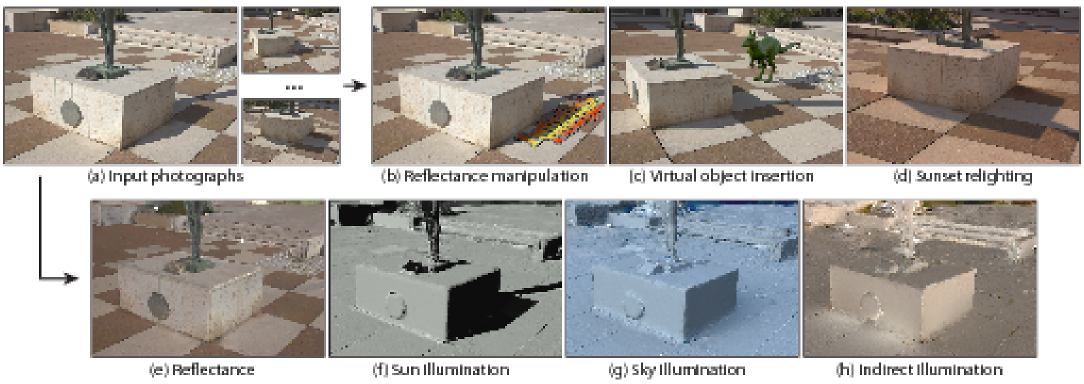


Figure 3. Starting from multiple views of the scene (a), our method decomposes photographs into four intrinsic layers — the reflectance (e), the illumination due to sun (f), the illumination due to sky (g) and the indirect illumination (h). Each layer can then be manipulated independently for advanced image editing applications (b-d).

5.1.2. Coherent Intrinsic Images from Photo Collections

Participants: Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

We propose a second method to compute intrinsic images in the presence of varying lighting conditions. Our method exploits the rich information provided by *multiple viewpoints and illuminations* in an image collection to process complex scenes without user assistance, nor precise and complete geometry. Such collections can be gathered from photo-sharing websites, or captured indoors with a light source which is moved around the scene.

We use multi-view stereo to automatically reconstruct 3D points and normals, from which we derive relationships between reflectance values at different locations, across multiple views, and consequently across different lighting conditions. In addition, we propose an optimization approach which enforces coherent reflectance in all views of a scene.

The resulting *coherent intrinsic images* enable image-based illumination transfer between photographs of the collection, as illustrated in Figure 4 .

This work is a collaboration with Frédo Durand (MIT) and Sylvain Paris (Adobe), and started with a visit of Pierre-Yves Laffont at MIT during Summer 2011. It has been published in the ACM Transactions on Graphics journal [19], and has been presented at SIGGRAPH Asia 2012.

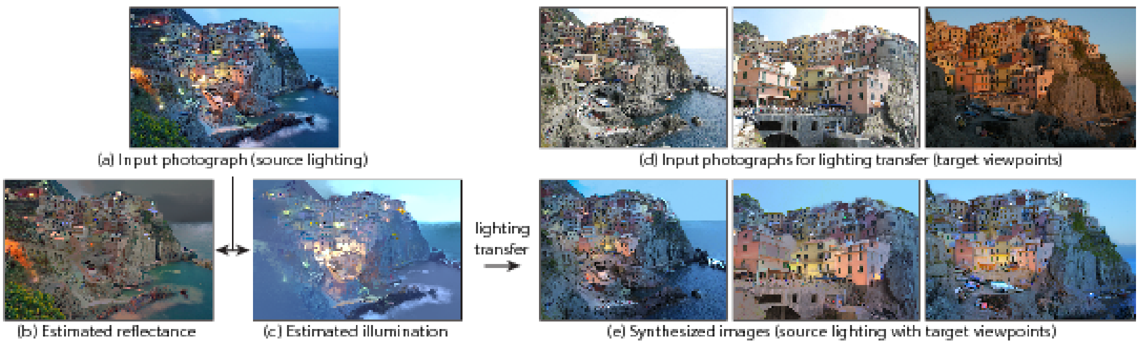


Figure 4. Our method automatically decomposes each image of a photo collection into reflectance and illumination (a-c). Transferring the illumination layer (c) to other viewpoints (d) yields synthetic images with novel viewpoint/lighting combinations (e).

5.1.3. Intrinsic Images by Clustering

Participant: Jorge Lopez Moreno.

Decomposing an input image into its intrinsic illumination and reflectance components is a long-standing ill-posed problem. We present a novel algorithm that requires no user strokes and works on a single image. Based on simple assumptions about its reflectance and luminance, we first find clusters of similar reflectance in the image, and build a linear system describing the connections and relations between them. Our assumptions are less restrictive than widely-adopted Retinex-based approaches, and can be further relaxed in conflicting situations. The resulting system is robust even in the presence of areas where our assumptions do not hold. We show a wide variety of results, including natural images, objects from the MIT dataset and texture images, along with several applications, proving the versatility of our method (see Figure 5).

This work is a collaboration with Elena Garces, Adolfo Munoz and Diego Gutierrez from University of Zaragoza (Spain). The work was published in an special issue of the journal Computer Graphics Forum and presented at the Eurographics Symposium on Rendering 2012 [16].

5.1.4. Relighting for Image Based Rendering

Participants: Sylvain Duchêne, Jorge Lopez Moreno, Stefan Popov, George Drettakis.

Figure 5. Decomposition by our method of the input image (left) into illumination (center) and reflectance (right) components.

Image-based rendering generates realistic virtual images from a small set of photographs. However, while current methods can simulate novel viewpoints from the input pictures, they cannot produce novel illumination conditions that differ from the lighting at the time of capture. The goal of this project is to provide such relighting capabilities. Our method first rely on multi-view stereo algorithms to estimate a coarse geometry of the scene. This geometry is often innacurate and incomplete. We complement it with image-based propagation algorithms that fill-in the missing data using the high-resolution input pictures. This combination of geometric and image-based cues allows us to generate plausible shadow motion and simulate novel sun directions.

5.1.5. Depth Synthesis and Local Warps for Plausible Image-based Navigation

Participants: Gaurav Chaurasia, Sylvain Duchêne, George Drettakis.

Modern multi-view stereo algorithms can estimate 3D geometry from a small set of unstrcutured photographs. However, the 3D reconstruction often fails on vegetation, vehicles and other complex geometry present in everyday urban scenes. We introduce a new Image-Based Rendering algorithm that is robust to unreliable geometry. Our algorithm segments the image into superpixels, *synthesizes* depth in superpixels with missing depth, warps them using a shape-preserving warp and blends them to create real-time plausible novel views for challenging target scenes, resulting in convincing immersive navigation experience.

This work is in collaboration with Dr. Olga Sorkine at ETH Zürich. and has been submitted to ACM Transactions on Graphics.

5.1.6. Perception of Slant for Image-Based Rendering

Participants: Christian Richardt, Peter Vangorp, George Drettakis.

Image-based rendering can create images with a high level of realism using simple geometry. However, as soon as the viewer moves away from the correct viewpoint, the image appears deformed. This work investigates the parameters which influence the perception of these image deformations. We propose a novel model of slant perception, which we validate using psychophysical experiments.

This work is a collaboration with Peter Vangorp at MPI Informatik, and Emily Cooper and Martin Banks from the University of California, Berkeley; in the context of the Associate Team CRISP (see also Section 7.3.1.1).

5.1.7. Lightfield Editing

Participant: Adrien Bousseau.

Lightfields capture multiple nearby views of a scene and are consolidating themselves as the successors of conventional photographs. As the field grows and evolves, the need for tools to process and manipulate lightfields arises. However, traditional image manipulation software such as Adobe Photoshop are designed to handle single views and their interfaces cannot cope with multiple views coherently. In this work we evaluate different user interface designs for lightfield editing. Our interfaces differ mainly in the way depth is presented to the user and build uppon different depth perception cues.

This work is a collaboration with Adrian Jarabo, Belen Masia and Diego Gutierrez from Universidad de Zaragoza and Fabio Pellacini from Sapienza Università di Roma.

5.1.8. Example-Based Fractured Appearance

Participants: Carles Bosch, George Drettakis.

A common weathering effect is the appearance of cracks due to material fractures. Previous exemplar-based aging and weathering methods have either reused images or sought to replicate observed patterns exactly. We propose an approach to exemplar-based modeling that creates weathered patterns by matching the statistics of fracture patterns in a photograph. We conducted a user study to determine which statistics are correlated to visual similarity and how they are perceived by the user. We describe a physically-based fracture model capable of producing similar crack patterns at interactive rates and an optimization method to determine its parameters based on key statistics of the exemplar. Our approach is able to produce a variety of fracture effects from simple crack photographs at interactive rates, as shown in Figure 6 .

Figure 6. Application of our example-based fracturing method on different scenes. Photographs of input fracture patterns are shown in the insets.

This work is a collaboration with Loeiz Glondu, Maud Marchal and George Dumont from IRISA-INSA/Inria Rennes - Bretagne Atlantique, Lien Muguercia from the University of Girona, and Holly Rushmeier from Yale University. The work was published in the Computer Graphics Forum journal and presented at the 23rd Eurographics Symposium on Rendering [17].

5.1.9. Real-Time Rendering of Rough Refraction

Participant: Adrien Bousseau.

We propose an algorithm to render objects made of transparent materials with rough surfaces in real-time, under all-frequency distant illumination. Rough surfaces cause wide scattering as light enters and exits objects, which significantly complicates the rendering of such materials. We present two contributions to approximate the successive scattering events at interfaces, due to rough refraction: First, an approximation of the Bidirectional Transmittance Distribution Function (BTDF), using spherical Gaussians, suitable for real-time estimation of environment lighting using pre-convolution; second, a combination of cone tracing and macro-geometry filtering to efficiently integrate the scattered rays at the exiting interface of the object. We demonstrate the quality of our approximation by comparison against stochastic ray-tracing (see Figure 7).

Furthermore we propose two extensions to our method for supporting spatially varying roughness on object surfaces and local lighting for thin objects.

(a) Ground truth

(b) Our method

Figure 7. Compared to an expensive ray-traced reference (a), our method produces plausible results in real-time (b).

This work is a collaboration with Charles De Rousiers , Kartic Subr, Nicolas Holzschuch from Inria Grenoble, and Ravi Ramamoorthi from UC Berkeley in the context of the Associate Team CRISP (see also Section 7.3.1.1). A paper describing the method was published in the IEEE Transactions on Visualization and Computer Graphics journal [14].

5.1.10. Gabor Noise by Example

Participants: Ares Lagae, George Drettakis.

Procedural noise is a fundamental tool in Computer Graphics. However, designing noise patterns is hard. In this project, we propose *Gabor noise by example*, a method to estimate the parameters of bandwidth-quantized Gabor noise, a procedural noise function that can generate noise with an arbitrary power spectrum, from exemplar Gaussian textures, a class of textures that is completely characterized by their power spectrum (see Figure 8).

More specifically, we introduce (i) bandwidth-quantized Gabor noise, a generalization of Gabor noise to arbitrary power spectra that enables robust parameter estimation and efficient procedural evaluation; (ii) a robust parameter estimation technique for quantized-bandwidth Gabor noise, that automatically decomposes the noisy power spectrum estimate of an exemplar into a sparse sum of Gaussians using non-negative basis pursuit denoising; and (iii) an efficient procedural evaluation scheme for bandwidth-quantized Gabor noise, that uses multi-grid evaluation and importance sampling of the kernel parameters. Gabor noise by example preserves the traditional advantages of procedural noise, including a compact representation and a fast on-the-fly evaluation, and is mathematically well-founded.

This work is a collaboration with Bruno Galerne from MAP5, Université Paris Descartes and CNRS, Sorbonne Paris Cité; Ares Lagae from KU Leuven; and Sylvain Lefebvre from the ALICE project team, Inria Nancy - Grand Est. This work was presented at SIGGRAPH 2012 and published in ACM Transactions on Graphics [15].

5.1.11. Structured Gabor noise

Participants: Gaurav Chaurasia, Ares Lagae, George Drettakis.

Current procedural noise synthesis techniques [15] are limited to Gaussian random field textures. This project aims to generalize procedural noise to a broader class of structured textures.

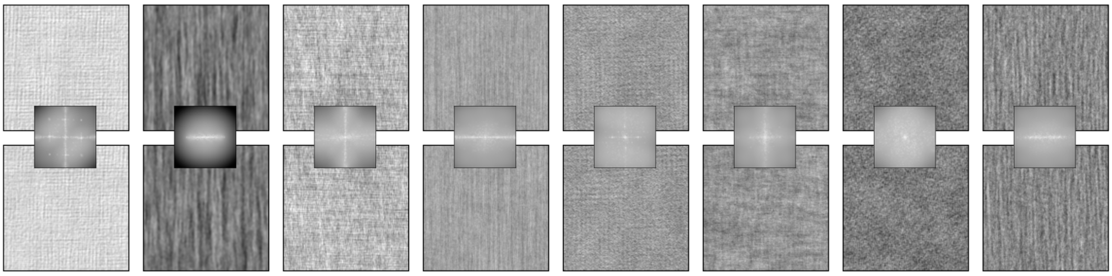


Figure 8. Gabor noise by example is a method to estimate the parameters of bandwidth-quantized Gabor noise, a procedural noise function that can generate noise with an arbitrary power spectrum, from exemplar Gaussian textures, a class of textures that is completely characterized by their power spectrum. (row 1) Gaussian texture. (row 2) Procedural noise. (insets) Estimated power spectrum.

This work is in collaboration with Dr. Ares Lagae (Katholieke Universiteit Leuven, Belgium), Dr. Bruno Galerne (Université Paris Descartes) and Prof. Ravi Ramamoorthi (UC Berkeley), in the contect of the Associate Team CRISP (Section 7.3.1.1).

5.1.12. Gloss Perception in Painterly and Cartoon Rendering

Participant: Adrien Bousseau.

Depictions with traditional media such as painting and drawing represent scene content in a stylized manner. It is unclear however how well stylized images depict scene properties like shape, material and lighting. In this project, we use non photorealistic rendering algorithms to evaluate how stylization alters the perception of gloss (see Figure 9). Our study reveals a compression of the range of representable gloss in stylized images so that shiny materials appear more diffuse in painterly rendering, while diffuse materials appear shinier in cartoon images.

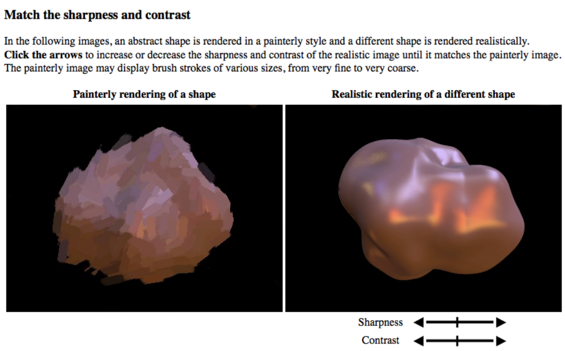


Figure 9. The experimental task used for studying gloss perception in stylized images.

From our measurements we estimate the function that maps realistic gloss parameters to their perception in a stylized rendering. This mapping allows users of NPR algorithms to predict the perception of gloss in their images. The inverse of this function exaggerates gloss properties to make the contrast between materials in a stylized image more faithful. We have conducted our experiment both in a lab and on a crowdsourcing website. While crowdsourcing allows us to quickly design our pilot study, a lab experiment provides more control on how subjects perform the task. We provide a detailed comparison of the results obtained with the two approaches and discuss their advantages and drawbacks for studies like ours.

This work is a collaboration with James O'Shea, Ravi Ramamoorthi and Maneesh Agrawala from UC Berkeley in the context of the Associate Team CRISP (see also Section 7.3.1.1) and Frédo Durand from MIT. It will be published in ACM Transactions on Graphics 2013 [12] (in press).

5.2. Interaction and Design for Audiovisual Virtual Environments

5.2.1. Auditory-visual integration of emotional signals in a virtual environment for cynophobia

Participants: Emmanuelle Chapoulie, Adrien David, Rachid Guerchouche, George Drettakis.

Cynophobia (dog phobia) has both visual and auditory relevant components. In order to investigate the efficacy of virtual reality exposure-based treatment for cynophobia, we studied the efficiency of auditory-visual environments in generating presence and emotion. We conducted an evaluation test with healthy participants sensitive to cynophobia in order to assess the capacity of auditory-visual virtual environments to generate fear reactions. Our application involves both high fidelity visual stimulation displayed in an immersive space and 3D sound. This specificity enables us to present and spatially manipulate fearful stimuli in the auditory modality, the visual modality and both.

We conducted a study where participants were presented with virtual dogs in realistic environments. Dogs were presented in a progressive manner, from unimodal and static to audiovisual and dynamic. Participants were also submitted a Behavioral Assessment Test at the beginning and end of the experiment where they were presented a virtual dog walking towards them step by step until it was extremely close. Finally, they completed several questionnaires and were asked to comment on their experience. The participants reported higher anxiety levels in response to auditory-visual stimuli compared to unimodal stimuli. Our results strongly suggest that manipulating auditory-visual integration might be a good way to modulate affective reactions and that auditory-visual VR are a promising tool for the treatment of cynophobia.

This work is a collaboration with Marine TAFFOU and Isabelle VIAUD-DELMON from IRCAM, in the context of ARC NIEVE (see also Section 7.1.4). The work was published in the Annual Review of Cybertherapy and Telemedicine in 2012.

5.2.2. Procedural audio modeling for particle-based environmental effects

Participants: Charles Verron, George Drettakis.

In this project we proposed a sound synthesizer dedicated to particle-based environmental effects, for use in interactive virtual environments. The synthesis engine is based on five physically-inspired basic elements which we call sound atoms, that can be parameterized and stochastically distributed in time and space. Based on this set of atomic elements, models are presented for reproducing several environmental sound sources. Compared to pre-recorded sound samples, procedural synthesis provides extra flexibility to manipulate and control the sound source properties with physically-inspired parameters. The controls are used to simultaneously modify particle-based graphical models, resulting in synchronous audio/graphics environmental effects. The approach is illustrated with three models, that are commonly used in video games: fire, wind, and rain. The physically-inspired controls simultaneously drive graphical parameters (e.g., distribution of particles, average particles velocity) and sound parameters (e.g., distribution of sound atoms, spectral modifications) as illustrated on Figure 10 for fire. The joint audio/graphics control results in a tightly-coupled interaction between the two modalities that enhances the naturalness of the scene.

The work was presented at the 133rd AES convention in October 2012 [23].

5.2.3. Perception of crowd sounds

Participants: Charles Verron, George Drettakis.

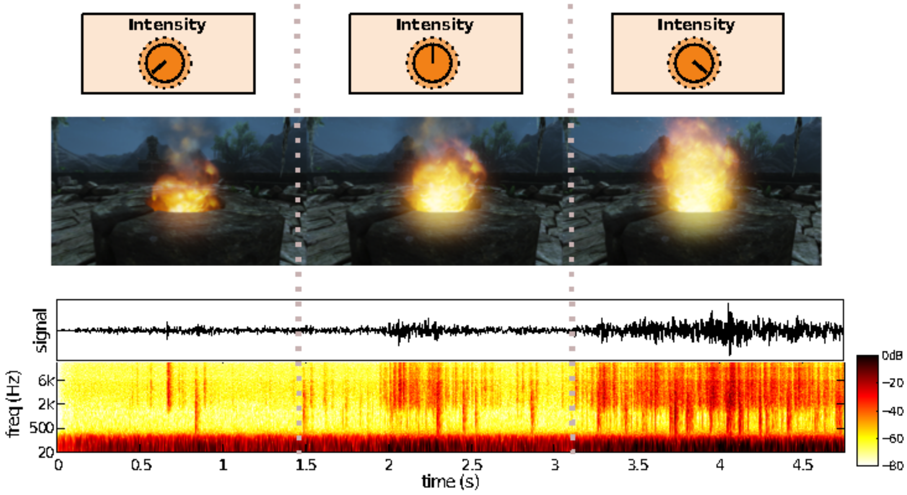


Figure 10. Audio/graphics high-level control of a fire. The control Intensity changes the rate and gain of noisy impacts, and the combustion noise of the fire sound model. Simultaneously, it controls the flame/smoke particle spawn rate for the graphics simulation.

Simulating realistic crowd scenes is an important challenge for virtual reality and games. Motion capture techniques allow to reproduce efficiently characters that look, move and sound realistic in virtual environments. However a huge amount of data is required to ensure that all agents behave differently in a big crowd. A common approach to solve this issue is to “clone" the same appearance, motion or sound several times, which can lead to perceived repetitions and break the realism of the scene. In this study we further investigate our perception of crowd scenes. Using a database of motions and sounds captured for 40 actors, along with a database of 40 different appearance templates, we propose an experimental framework to evaluate the perceptual degradations caused by clones. A particular attention is given to evaluate the influence of appearance, motion and sound, either separately or in multimodal conditions. This study aims at providing useful insights on our perception of crowd scenes, and guidelines to designers in order to reduce the amount of resources to produce convincing crowd scenes.

This ongoing project is a collaboration between Inria, CNRS-LMA (Marseille, France) and Trinity College (Dublin, Ireland).

5.2.4. Walking in a Cube: Novel Metaphors for Safely Navigating Large Virtual Environments in Restricted Real Workspaces

Participants: Peter Vangorp, Emmanuelle Chapoulie, George Drettakis.

Immersive spaces such as 4-sided displays with stereo viewing and high-quality tracking provide a very engaging and realistic virtual experience. However, walking is inherently limited by the restricted physical space, both due to the screens (limited translation) and the missing back screen (limited rotation). Locomotion techniques for such restricted workspaces should satisfy three concurrent goals: keep the user safe from reaching the translational and rotational boundaries; increase the amount of real walking; and finally, provide a more enjoyable and ecological interaction paradigm compared to traditional controller-based approaches.

We have proposed three novel locomotion techniques that attempt to satisfy these goals in innovative ways. We constrain traditional Wand locomotion by turning off the Wand controls for directions that can be reached by real walking instead, and we display warning signs when the user approaches the limits of the real workspace (Figure 11 (a)). We also extend the Magic Barrier Tape paradigm with “blinders” to avoid rotation towards the missing back screen (Figure 11 (b)). Finally, we introduce the "Virtual Companion", which uses a small bird to guide the user through virtual environments larger than the physical space (Figure 11 (c,d)).

We evaluate the three new techniques through a user study with travel-to-target and path following tasks. The study provides insight into the relative strengths of each new technique for the three aforementioned goals. Specifically, if speed and accuracy are paramount, traditional controller interfaces augmented with our novel warning techniques may be more appropriate; if physical walking is more important, two of our paradigms, the extended Magic Barrier Tape and the Constrained Wand, should be preferred; and finally, fun and ecological criteria would favor the Virtual Companion.

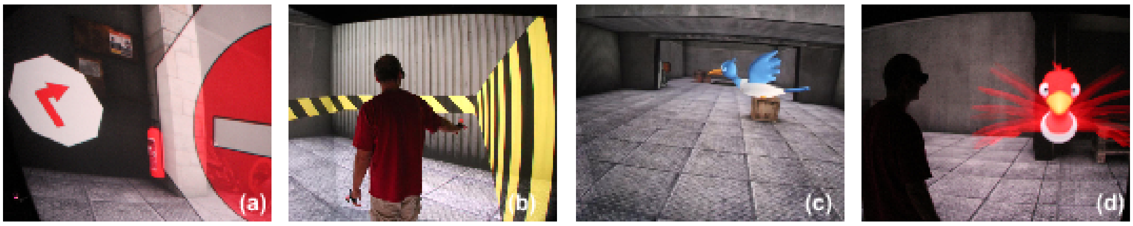


Figure 11. Screenshots illustrating the three novel locomotion techniques. From left to right: (a) Constrained Wand and signs: the "no-way" and "turn right" signs. (b) Extended Magic Barrier Tape: the tape and blinders. (c,d) Virtual Companion: the bird in "rest mode" (c) and "protection mode" (d).

This work is a collaboration with Gabriel Cirio, Maud Marchal and Anatole Lécuyer (VR4I project team, IRISA-INSA/Inria Rennes - Bretagne Atlantique) in the context of ARC NIEVE (see Section 7.1.4). The work was published in the special issue of the journal IEEE Transactions on Visualization and Computer Graphics (TVCG) [13], and presented at the IEEE Virtual Reality conference 2012.

5.2.5. Natural Gesture-based Interaction for Complex Tasks in an Immersive Cube

Participants: Emmanuelle Chapoulie, Jean-Christophe Lombardo, George Drettakis.

We present a solution for natural gesture interaction in an immersive cube in which users can manipulate objects with fingers of both hands in a close-to-natural manner for moderately complex, general purpose tasks. To do this, we develop a solution using finger tracking coupled with a real-time physics engine, combined with a comprehensive approach for hand gestures, which is robust to tracker noise and simulation instabilities. To determine if our natural gestures are a feasible interface in an immersive cube, we perform an exploratory study for tasks involving the user walking in the cube while performing complex manipulations such as balancing objects. We compare gestures to a traditional 6-DOF Wand, and we also compare both gestures and Wand with the same task, faithfully reproduced in the real world. Users are also asked to perform a free task, allowing us to observe their perceived level of presence in the scene. Our results show that our robust approach provides a feasible natural gesture interface for immersive cube-like environments and is perceived by users as being closer to the real experience compared to the Wand.

This work is a collaboration with Evanthia Dimara and Maria Roussou from the University of Athens and with Maud Marchal from IRISA-INSA/Inria Rennes - Bretagne Atlantique. The work has been submitted to 3DUI 2013.

5.2.6. CrossShade: Shading Concept Sketches Using Cross-Section Curves

Participant: Adrien Bousseau.

We facilitate the creation of 3D-looking shaded production drawings from concept sketches. The key to our approach is a class of commonly used construction curves known as cross-sections, that function as an aid to both sketch creation and viewer understanding of the depicted 3D shape. In particular, intersections of these curves, or cross-hairs, convey valuable 3D information, that viewers compose into a mental model of the overall sketch. We use the artist-drawn cross-sections to automatically infer the 3D normals across the sketch, enabling 3D-like rendering (see Figure 12).

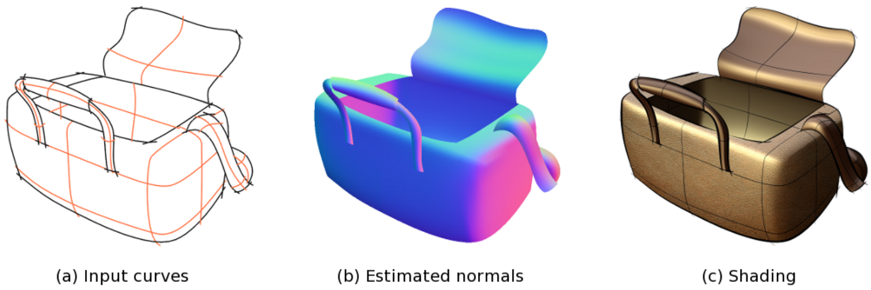


Figure 12. Concept sketches (a) frequently use cross-sections (drawn in orange) to convey 3D shape with just a handful of lines. We derive the mathematical properties of cross-section curves and leverage them to automatically estimate surface normals across the drawn objects (b). The resulting normal field allow users to shade the objects using a variety of shading styles and setups (c).

The technical contribution of our work is twofold. First, we distill artistic guidelines for drawing cross-sections and insights from perception literature to introduce an explicit mathematical formulation of the relationships between cross-section curves and the ge- ometry they aim to convey. We then use these relationships to develop an algorithm for estimating a normal field from cross-section curve networks and other curves present in concept sketches. We validate our formulation and algorithm through a user study and a ground truth normal comparison. These contributions enable us to shade a wide range of concept sketches with a variety of rendering styles.

This work is a collaboration with Cloud Shao and Karan Singh from the University of Toronto and Alla Sheffer from the University of British Columbia. It has been published at ACM Transactions on Graphics, proceedings of the SIGGRAPH 2012 conference.

5.2.7. CrossShape

Participant: Adrien Bousseau.

We facilitate the automatic creation of surfaced 3D models from design sketches that employ a commonly drawn network of cross-section curves. Our previous method generates 3D renderings of input sketches by creating a 3D surface normal field that interpolates the sketched cross-sections. This normal field however, incorporates the inevitable inaccuracy of sketched curves, making it inappropriate for 3D surface construction.

Successful construction of the 3D surface perceived from sketches requires cross-section properties and other perceived curve relationships such as symmetry and parallelism, to be met precisely. We present a novel formulation where these geometric constraints are satisfied while minimizing the difference between the sketch and the 3D cross-sections projected on it. We validate our approach by producing accurate surface reconstructions of existing 3D models represented using a network of cross-sections as well on a variety of sketch input. Finally we illustrate our surfacing solution within an interactive sketch based modeling framework.

This work is a collaboration with James McCrae and Karan Singh from the University of Toronto and Xu Baoxuan and Alla Sheffer from the University of British Columbia.

5.2.8. Computer-assisted drawing

Participants: Emmanuel Iarussi, Adrien Bousseau.

A major challenge in drawing from observation is to trust what we *see* rather than what we *know*. Drawing books and tutorials provide simple techniques to gain consciousness of the shapes that we observe and their relationships. Common techniques include drawing simple geometrical shapes first – also known as *blocking in* – and checking for alignments and equal proportions. While very effective, these techniques are usually illustrated on few examples and it takes significant effort to generalize them to an arbitrary model. In addition, books and tutorials only contain static instructions and cannot provide feedback to people willing to practice drawing.

In this project, we develop an interactive drawing tool that assists users in their practice of common drawing techniques. Our *drawing assistant* helps users to draw from any model photograph and provides corrective feedback interactively.

This work is a collaboration with Theophanis Tsandilas from the InSitu project team, Inria Saclay - Ile de France, in the context of the ANR DRAO project (see Section 7.1.2).

5.2.9. Depicting materials in vector graphics

Participants: Jorge Lopez-Moreno, Adrien Bousseau, Stefan Popov, George Drettakis.

Vector drawing tools like Illustrator and InkScape enjoy great popularity in illustration and design because of their flexibility, directness and distinctive look. Within such tools, skillful artists depict convincing material and lighting effects using 2D vector primitives like gradients and paths. However, it takes significant expertise to convey plausible material appearance in vector drawings. Instead, novice users often fill-in regions with a constant color, sacrifying plausibility for simplicity. In this project we present the first vector drawing tool that automates the depiction of material appearance. Users can use our tool to either fill-in regions automatically, or to generate an initial set of vector primitives that they can refine at will.

This work is a collaboration with Maneesh Agrawala from the University of Berkeley in the context of the Associate Team CRISP (see Section 7.3.1.1).

5.2.10. Gradient Art: Creation and Vectorization (survey)

Participant: Adrien Bousseau.

We survey the main two categories of methods for producing vector gradients. One is mainly interested in converting existing photographs into dense vector representations. By vector it is meant that one can zoom infinitely inside images, and that control values do not have to lie onto a grid but must represent subtle color gradients found in input images. The other category is tailored to the creation of images from scratch, using a sparse set of vector primitives. In this case, we still have the infinite zoom property, but also an advanced model of how space should be filled in-between primitives, since there is no input photograph to rely on. These two categories are actually extreme cases, and seem to exclude each other: a dense representation is difficult to manipulate, especially when one wants to modify topology; a sparse representation is hardly adapted to photo vectorization, especially in the presence of texture. Very few methods lie in the middle, and the ones that do require user assistance.

We published our survey in the book *Image and Video based Artistic Stylization* [25] editied by Springer. The survey was written in Collaboration with Pascal Barla from the MANAO project team, Inria Bordeaux - Sud Ouest, in the context of the ANR DRAO project (see Section 7.1.2).

SEMAGRAMME Team

6. New Results

6.1. Syntax-Semantics Interface

6.1.1. Graph Rewriting

Bruno Guillaume and Guy Perrier have proposed a system for annotating the French Treebank with semantic dependencies [12], [14]. This system (Synsem_FTB) is based on Graph Rewriting. Graph Rewriting is a framework which is well-suited for syntax-semantic interface because it allows for a modular development of large systems. Each modelled linguistic phenomenon is described by a small set of local rewriting rules. The whole transformation is then described by a sequence of modules to apply successively to the input structure. Another benefit of the Graph Rewriting formalism is that it handles the ambiguity in a natural way with the use of non confluent rewriting systems.

The Synsem_FTB system produces a semantic annotation in the framework of DMRS starting from an annotation with surface syntactic dependencies. It contains 34 modules that can be split in two main parts; the first part produces a deep syntax annotation of the input and the second one rewrites deep syntax to semantics.

With respect to previous works, the system of rewriting rules itself has been improved: it has a larger coverage (causative constructions, rising verbs, ...) and the order between modules has been studied in a more systematic way.

The rewriting calculus has been enriched on two points: the use of rules to make a link with lexicons, especially with the lexicon of verbs Dicovalence, and the introduction of filters to discard inconsistent annotations at some computation steps.

This system has been experimented on the whole French Treebank with the Grew software, which implements the used rewriting calculus.

6.1.2. Passive Sentences

Chris Blom, Philippe de Groote, Yoad Winter, and Joost Zwarts have proposed a unified syntactic-semantic account of passive sentences and sentences with an unspecified object [18]. For both constructions, they use *option types* for introducing implicit arguments into the syntactic-semantic categorial mechanism. They show the advantages of this approach over previous proposals in the domains of scope and unaccusatives. Unlike pure syntactic treatments, option types immediately derive the obligatory narrow scope of existential quantification over an implicit argument's slot. Unlike purely semantic, event-based treatments, their solution naturally accounts for syntactic contrasts between passives and unaccusatives.

6.1.3. Intensionalization

Makoto Kanazawa and Philippe de Groote have defined a general *intensionalization* procedure that turns an extensional semantics for a language into an intensionalized one that is capable of accommodating *truly intensional* lexical items without changing the compositional semantic rules [48]. They have proved some formal properties of this procedure and have clarified its relation to the procedure implicit in Montague's PTQ.

6.1.4. Plural

Sai Qian and Maxime Amblard have modeled the semantics of plurality in continuation semantics [13]. Two types of discourse antecedents formations, inherited from the classical treatment, namely summation and abstraction, are studied in detail. Solutions for each phenomenon are provided respectively by introducing two new functions `Sum` and `Abs`, for obtaining the semantic interpretations.

6.2. Discourse Dynamics

In a joint work with a psycho-linguistist (Michel Musiol, ATILF) and a philosopher (Manuel Rebuschi, Archives Poincaré), are developing a formal analysis of pathological conversations involving schizophrenic speakers [16]. Such conversations give rise to manifest incongruities or ruptures that can be seen as mere contradictions by any “normal” speaker. Our analysis relies both on semantic and pragmatic features of conversation. We propose a SDRT-inspired [20] account of pathological conversations, and we apply it to two relevant excerpts. We conclude with a short discussion about the localization of inconsistencies by schizophrenics, either in semantics or in pragmatics, and its importance for our understanding of thought disorders.

SIROCCO Project-Team

6. New Results

6.1. Analysis and modeling for compact representation and navigation

3D modelling, multi-view plus depth videos, Layered depth images (LDI), 2D and 3D meshes, epitomes, image-based rendering, inpainting, view synthesis

6.1.1. Computational modelling of visual attention

Participants: Josselin Gautier, Olivier Le Meur, Zhi Liu.

6.1.1.1. Time-dependent saliency map

The study related to the deployment of visual attention in 2D and 3D has been completed in 2012. The purpose of this study was to investigate whether or not there is a difference between eye movements recorded while observers viewed natural images in 2D and 3D conditions. Results show that visual exploration in depth layer detection task is affected by the binocular disparity. In particular, participants tend to look first at closer areas just after the stimuli onset with the introduction of disparity, and then direct their gaze to more widespread locations. Based on these conclusions, a computational model of visual attention taking into account the temporal dimension has been designed. An Expectation-Maximisation (EM) algorithm has been used to infer the weight of different visual features (saliency, depth, center bias) over time. Results have been published in the journal Cognitive Computation.

A new study on a similar subject has started during the summer 2012. The purpose is again to investigate the influence of binocular disparity, scene complexity on visual scanpaths obtained in 2D and 3D viewing conditions. The main differences with the previous study are twofold. First, a new database of content has been designed. All parameters such as the amount of disparity are accurately mastered. Second is about the context of the study which deals with quality assessment of 3D video content.

6.1.1.2. Salient object detection

In 2012, Dr. Liu, who has joined the team in August for 2 years has started a study dealing with salient object detection. The goal is to extract automatically the most interesting object in an image or video sequence. The proposed approach is based on low-level visual features and extensively used a superpixel method. Starting from the superpixel representation of an image, the saliency measure of each superpixel is evaluated based on its global uniqueness and local contrasts with other superpixels. A saliency-directed region merging algorithm with a dynamic scale control scheme is then exploited to generate more meaningful regions. The region merging process is recorded using a Binary Partition Tree (BPT), in which each leaf node represents each superpixel and each non-leaf node represents each generated region during the region merging process. Finally, a node selection algorithm based on saliency density difference is used to select suitable nodes from BPT to form the salient object detection result. First experimental results on a public dataset (MSRA) are promising and demonstrate the effectiveness of the proposed approach.

6.1.2. Similarity metrics for image processing

Participants: Mounira Ebdelli, Christine Guillemot, Olivier Le Meur, Raul Martinez Noriega, Aline Roumy.

Several image processing problems addressed by the team (inpainting, loss concealment, super-resolution, denoising) require having patch objective similarity metrics as close as possible to ground truth visual similarity. The derivation of such metrics has been investigated along several directions. First, a performance analysis of the most used fidelity metrics (SSD, SSIM, two SSD-weighted Battacharya metrics) has been carried out to assess the perceptual similarities between patches. A statistical analysis of subjective tests has shown that some of these metrics (the SSD-weighted Battacharya) are more suitable than others to respect human decisions in terms of patch similarities. This conclusion has been confirmed with the results of Non Local means (NL-means) denoising algorithm which are highly sensitive to the used similarity metrics. The value of each pixel p in the blurred image is updated using a weighted average of the collocated pixels values in the most similar patches to the block centered on p. We show that SSD, which is the most used similarity metric, is not necessary the best correlated with the perceptual criteria.

Greedy algorithms for inpainting are based on the assumption of self-similarity within an image. A patch located on the boundary of the hole to be filled in, contains a known part and an unknown part. The known part is used to select other (completely known) patches and called exemplars. Then, these exemplars are used to reconstruct the unknown part of the patch being processed. Such an approach faces two main problems, decision of filling-in order and selection of good exemplars from which the missing region is synthesized. In [29], we proposed an algorithm that tackles these problems with improvements in the preservation of linear edges, and reduction of error propagation compared to well-known algorithms from the literature. Our improvement in the filling-in order is based on a combination of priority terms, previously defined, that better encourages the early synthesis of linear structures. The second contribution helps reducing the error propagation thanks to a better detection of outliers from the candidate patches carried. This is obtained with a new metric based on the Hellinger distance between the patches that incorporates the whole information of the candidate patches.

6.1.3. Epitome-based image representation

Participants: Safa Cherigui, Christine Guillemot.

This work is carried out in collaboration with Technicolor (D. Thoreau, Ph. Guillotel, P. Perez) and aims at designing a compresion algorithm based on the concept of epitomes. An epitome is a condensed representation of an image (or a video) signal containing the essence of the textural properties of this image. Different forms of epitomes have been proposed in the literature, such as a patch-based probability model learned either from still image patches or from space-time texture cubes taken from the input video. These probability models together with appropriate inference algorithms, are useful for content analysis inpainting or super-resolution. Another family of approaches makes use of computer vision techniques, like the KLT tracking algorithm, in order to recover self similarities within and across images. In parallel, another type of approach consists in extracting epitome-like signatures from images using sparse coding and dictionary learning.

The method developed aims at tracking self-similarities within an image using a block matching (BM) algorithm. The epitome is constructed from disjoint pieces of texture (“epitome charts”) taken from the original image and a transform map which contains translational parameters. Those parameters keep track of the correspondences between each block of the input image and a block of the epitome. An Intra image compression scheme based on the epitome has been developed showing a rate saving of up to 12% on some images, including the rate cost of the epitome texture and of the transform map. The entire image can be reconstructed from the epitome texture with the help of the transform map.

6.2. Rendering, inpainting and super-resolution

image-based rendering, inpainting, view synthesis, super-resolution

6.2.1. Joint projection/filling method for virtual view synthesis

Participants: Christine Guillemot, Fabien Racapé.

This study is carried out in collaboration with INSA/IETR (Luce Morin). Associated with a view synthesis method, a multi-view plus depth video allows the generation of virtual views of the scene from any viewpoint. State-of-the-art synthesizers use Depth-Image-Based Rendering (DIBR) techniques based on warping equations, which project a reference view onto a virtual viewpoint. In classical DIBR schemes, the rendering proceeds in several distinct steps, each one designed to solve a specific problem. First, the depth map is warped onto the virtual viewpoint and filtered with a median filter. The filtered depth map is then used in a backward warping of the virtual view (as illustrated in FIg.1). The resulting depth map is inpainted, to fill in disocclusion areas. Finally, this complete depth map is used by a depth-aided inpainting algorithm to fill in disocclusions in the color map. However, all these steps are inter-dependent, and errors introduced by each one are amplified by the following one, producing annoying artifacts, as shown in Fig. 2 -(c).

The proposed Joint Projection Filling (JPF) method performs forward projection, using connectivity information to fill in disocclusions in a single step. Applied on the depth map warping, JPF enables a depth-aided inpainting of color disocclusions after backward projection, as shown in Fig. 1 . Fig. 2 -(e) presents a resulting synthesis which contains less artifacts.

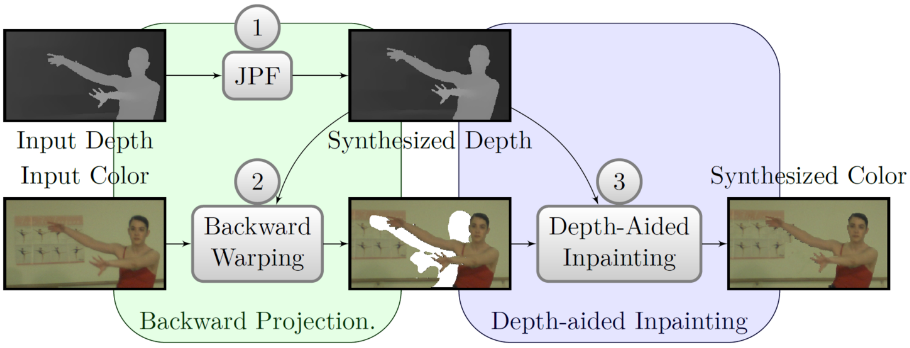


Figure 1. Virtual view generation chain, based on Joint Filling Projection. The depth map is jointly warped and inpainted. Depth-aided inpainting can be then used on disoccluded areas.

In the context of multi-view plus depth video coding (3D-HEVC standardization), inter view coding tools are added in the vein of temporal inter frame coding. We have tested our method as a projection tool for View Synthesis Prediction (VSP). However, the 3D-HEVC common test conditions, limited to rectified views as input, restrict the possible gains induced by efficient projection tools. Moreover, JPF outperforms other methods in synthesizing disoccluded areas with a good visual quality where VSP tools are not selected by MSE-based decision. JPF remains an efficient tool for extrapolating multi-view plus depth content with a minimum of artifacts on disoccluded areas.

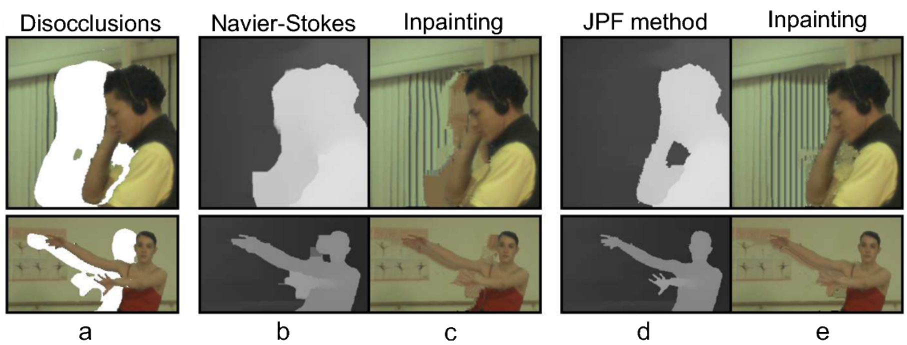


Figure 2. Disocclusion filling. (a) warped image before inpainting. Depth map inpainting: Navier-Stokes (b), JPF (d). Resulting depth-aided inpainting: Navier-Stokes (c), JPF (e).

6.2.2. Image inpainting using neighbor embedding and super-resolution

Participants: Mounira Ebdelli, Christine Guillemot, Olivier Le Meur.

Inpainting methods play an important role in a wide range of applications. Removing text and advertisements (such as logos), removing undesired objects, noise reduction and image reconstruction from incomplete data are the key applications of inpainting methods. Algorithms can be classified into two categories: PDE (Partial Derivative Equation)-based schemes and examplar-based schemes. The former uses diffusion schemes in order to propagate structures in a given direction. Their drawback is the introduction of blur due to diffusion. The latter relies on the sampling and the copying of texture from the known parts of the picture.

Image inpainting is a problem of texture synthesis. Given observations, or known samples in a spatial neighborhood, the goal is to estimate unknown samples of the patch to be filled in. Novel inpainting methods have been developed in the team along complementary directions: 1/- considering new priority functions exploiting the structure within the patch for defining the patch processing order; 2/- investigating various neighbor embedding techniques for estimating the unknown pixels; 3/- considering a coarse to fine multi-resolution approach where a low resolution version of the input image is first inpainted, this first step being followed by a a super-resolution based enhancement of the image.

Priority functions: Different priority functions using structure tensors and edge based information have been considered and their advantage over classical functions projecting isophote directions on the normal to the front line has been demonstrated.

Neighbor-embedding based inpainting: Examplar-based inpainting algorithms using various neighbor embedding techniques (LLE, LLE-LDNR, NMF with various solvers) have been developed. The methods have been shown to enhance the quality of inpainted images when compared to classical examplar-based solutions using simple template matching techniques to estimate the missing pixels, or similarity weights (NLM) (see Fig. 3).

Figure 3. Inpainting results: Original image; Mask of the image to be inpainted; Inpainting results with examplar-based inpainting using similarity weights; Inpainting results with neighbor embedding (LLE-LDNR).

Super-resolution aided inpainting: A novel super-resolution aided inpainting approach has been introduced which consists in first inpainting a coarse version of the input image and then in a second step, using a hierarchical super-resolution algorithm, to recover the native resolution [28]. However, to be less sensitive to the setting of the inpainting methods, the low-resolution input picture is inpainted several times with different settings. Results are efficiently combined with a loopy belief propagation. A super-resolution algorithm is then applied to recover the details. Experimental results in a context of image editing, texture synthesis and 3D view synthesis demonstrate the effectiveness of the proposed method. Fig.4 show texture synthesis results obtained with this approach.

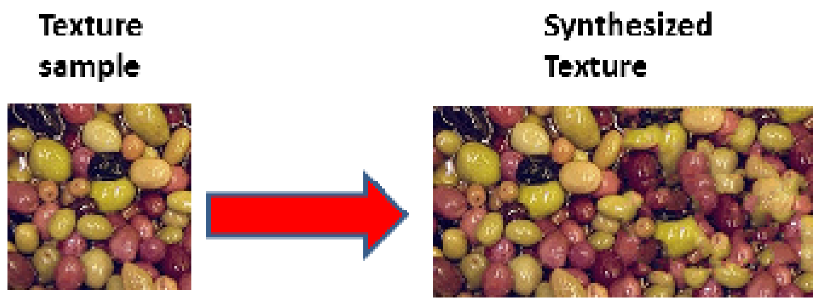


Figure 4. Texture synthesis results obtained with super-resolution aided inpainting.

6.3. Representation and compression of large volumes of visual data

sparse representations, data dimensionality reduction, compression, scalability, perceptual coding, rate-distortion theory

6.3.1. Multi-view plus depth video compression

Participants: Christine Guillemot, Thomas Guionnet, Laurent Guillo, Fabien Racapé.

Multi-view plus depth video content represent very large volumes of input data wich need to be compressed for storage and tranmission to the rendering device. The huge amount of data contained in multi-view sequences indeed motivates the design of efficient representation and compression algorithms. In collaboration with INSA/IETR (Luce Morin), we have studied layered depth image (LDI) and layered depth video (LDV) representations as a possible compact representation format of multi-view video plus depth data. LDI give compact represention of 3D objects, which can be efficiently used for photo-realistic image-based rendering (IBR) of different scene viewpoints, even with complex scene geometry. The LDI extends the 2D+Z representation, but instead of representing the scene with an array of depth pixels (pixel color with associated depth values), each position in the array may store several depth pixels, organised into layers. A novel object-based LDI representation which is more tolerant to compression artifacts, as well as being compatible with fast mesh-based rendering techniques has been developped.

The team has also studied motion vector prediction in the context of HEVC-compatible Multi-view plus depth (MVD) video compression. The HEVC compatible MVD compression solution implements a 6 candidate vector list for merge and skip modes. As part of the 3D video encoding, an inter–view motion vector predictor is added at the first position of this list. Our works show that this new list can be improved in optimizing the order of the candidates and in adding two more relevant candidates. When a merge or a skip mode is selected, a merge index is written in the bitstream. This index is first binarized using a unary code, then encoded with the CABAC. A CABAC context is dedicated to the first bin of the unary coded index while the remaining bins are considered as equiprobable. This strategy is efficient as long as the candidate list is ordered by decreasing index occurrence probability. However, this is not always the case when the inter-view motion vector predictor is added. To dynamically determine which candidate is the most probable, a merge index histogram is computed on the fly at the encoder and decoder side. Thus a conversion table can be calculated. It allows deriving the merge index to encode given the actual index in the list, and conversely, the actual index in the list given a decoded index. When using dynamic merge index, index re-allocation can happen at any time. Statistics of the first bin, which is encoded with CABAC, are modified. That is why a set of 6, one for each possible permutation of indexes, CABAC contexts dedicated to the first bin is defined. A bit rate gain of 0.1% for side views is obtained with no added complexity. These results are improved and reach 0.4% when additional CABAC contexts are used to take into account also the first three bins.

Candidates added by default in the merge list are not always the most relevant. As part of 3D video encoding using multiple rectified views, having a fine horizontal adjustment might be meaningful for efficient disparity compensated prediction. Therefore, we have proposed to replace some candidates in the merge list with candidates pointing to the base view and shifted by the horizontal offsets +4 and -4. To do so, the merge list is scanned to get among the first four candidates the first disparity compensated candidate. Once this vector found, the +4 and -4 offsets are added to its horizontal component and the two resulting vectors are inserted in the list two positions further if there is still room just after otherwise. With this improvement, a bit rate gain of 0.3% for side views is obtained with no added complexity.

6.3.2. Diffusion-based depth maps coding

Participants: Josselin Gautier, Olivier Le Meur.

A novel approach to compress depth map has been developed [26]. The proposed method exploits the intrinsic depth maps properties. Depth images indeed represent the scene surface and are characterized by areas of smoothly varying grey levels separated by sharp edges at the position of object boundaries. Preserving these characteristics is important to enable high quality view rendering at the receiver side. The proposed algorithm proceeds in three steps: the edges at object boundaries are first detected using a Sobel operator. The positions of the edges are encoded using the JBIG algorithm. The luminance values of the pixels along the edges are then encoded using an optimized path encoder. The decoder runs a fast diffusion-based inpainting algorithm which fills in the unknown pixels within the objects by starting from their boundaries.

6.3.3. Neighbor embedding for image prediction

Participants: Safa Cherigui, Christine Guillemot.

The problem of texture prediction can be regarded as a problem of texture synthesis. Given observations, or known samples in a spatial neighborhood, the goal is to estimate unknown samples of the block to be predicted. We have in 2010 and 2011 developed texture prediction methods as well as inpainting algorithms based on neighbor embedding techniques which come from the area of data dimensionality reduction [18], [31], [27]. The methods which we have more particularly considered are Locally Linear Embedding (LLE), LLE with Low-dimensional neighborhood representation (LDNR), and Non-negative Matrix Factorization (NMF) using various solvers.

The first step in the developed methods consists in searching, within the known part of the image, for the K nearest (KNN) patches to the set of known samples in the neighborhood of the block to be predicted (or of samples to be estimated in the context of inpainting). In a prediction (compression) context, in order for the decoder to proceed similarly, the K nearest neighbors are found by computing distances between the known pixels in a causal neighborhood (called template) of the input block and the co-located pixels in candidate patches taken from a causal window. Similarly, the weights used for the linear approximation are computed in order to best approximate the template pixels. Although efficient, these methods suffer from limitations when the template and the block to be predicted are not correlated, e.g. in non homogenous texture areas. To cope with these limitations, we have developed new image prediction methods based on neighbor embedding techniques in which the K-NN search is done in two steps and aided, at the decoder, by a block correspondence map, hence the name Map-Aided Neighbor Embedding (MANE) method. Another optimized variant of this approach, called oMANE method, has also been introduced. The resulting prediction methods are shown to bring significant Rate-Distortion (RD) performance improvements when compared to H.264 Intra prediction modes (up to 44.75%) [13]. Figure 5 illustrates the prediction quality obtained with different neighbor embedding methods, as well as the encoder selection rate of the oMANE-based prediction mode. This method has been presented at the IEEE International ICIP conference and the paper has been selected among the 11 finalists (out of 500 student papers) for a best student paper award.

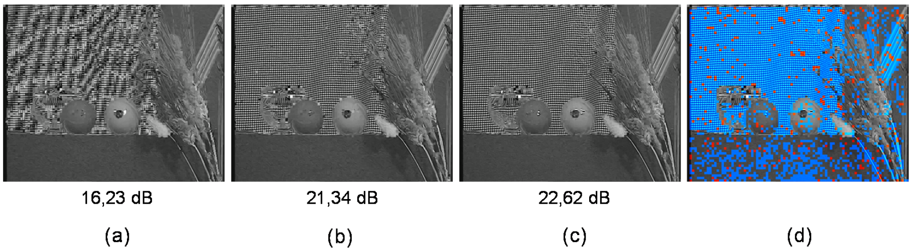


Figure 5. Spatial prediction for "Snook" with modes dynamically chosen according to a RD criterion with (a) H.264 Intra modes (High Profile), (b) LLE-based prediction, (c) Hybrid LLE-oMANE prediction and (d) selection rate of the two modes: LLE (red) and oMANE (blue).

6.3.4. Generalized lifting for video compression

Participants: Christine Guillemot, Bihong Huang.

This research activity is carried out in collaboration with Orange labs (Felix Henry) and UPC (Philippe Salembier) in Barcelona. The objective is to design new algorithmic tools for efficient loosless and lossy compression using generalized lifting concepts. The generalized lifting is a framework which permits the creation of nonlinear and signal probability density function (pdf) dependent and adaptive transforms. The use of such adaptive transforms for efficient coding of different HEVC syntax elements is under study.

6.3.5. Dictionary learning methods for sparse coding of satellite images

Participants: Jeremy Aghaei Mazaheri, Christine Guillemot, Claude Labit.

In the context of the national partnership Inria-Astrium, we explore novel methods to encode sequences of satellite images with a high degree of restitution quality and with respect to usual constraints in the satellite images on-board codecs. In this study, a geostationary satellite is used for surveillance and takes sequences of images. Then these pictures are stabilized and have to be compressed on-board before being sent to earth. Each picture has a high resolution and so the rate without compression is very high (about 70 Gbits/sec) and the goal is to achieve a rate after compression of 600 Mbits/sec, that is a compression ratio more than 100. On earth, the pictures are decompressed with a high necessity of reconstruction quality, especially for moving areas, and visualized by photo-interpreters. That is why the compression algorithm requires here a deeper study.The first stage of this study is to develop dictionary learning methods for sparse representations and coding of the images. These representations are commonly used for denoising and more rarely for image compression.

Sparse representation of a signal consists in representing a signal $y \in \Re^n$ as a linear combination of columns, known as atoms, from a dictionary matrix. The dictionary $D \in \Re^{n \times K}$ is generally overcomplete and contains K atoms. The approximation of the signal can thus be written $y \approx Dx$ and is sparse because a small number of atoms of D are used in the representation, meaning that the vector x has only a few non-zero coefficients. The choice of the dictionary is important for the representation. A predetermined transform matrix, as overcomplete wavelets or DCT, can be chosen. Another option is to learn the dictionary from training signals to get a well adapted dictionary to the given set of training data. Previous studies demonstrated that dictionaries have the potential to outperform the predetermined ones. Various advanced dictionary learning schemes have been proposed in the literature, so that the dictionary used is well suited to the data at hand. The popular dictionary learning algorithms include the K-SVD, the Method of Optimal Directions (MOD), Sparse Orthonormal Transforms (SOT), and (Generalized) Principle Component Analysis (PCA).

Recently, the idea of giving relations between atoms of a dictionary appeared with tree-structured dictionaries. Hierarchical sparse coding uses this idea by organizing the atoms of the dictionary as a tree where each node corresponds to an atom. The atoms used for a signal representation are selected among a branch of the tree. The learning algorithm is an iteration of two steps: hierachical sparse coding using proximal methods and update of the entire dictionary. Even if it gives good results for denoising, the fact to consider the tree as a single dictionary makes it, in its current state, not well adapted to efficiently code the indices of the atoms to select when the dictionary becomes large. We introduce in this study a new method to learn a tree-structured dictionary offering good properties to code the indices of the selected atoms and to efficiently realize sparse coding. Besides, it is scalable in the sense that it can be used, once learned, for several sparsity constraints. We show experimentally that, for a high sparsity, this novel approach offers better rate-distortion performances than state-of-the-art "flat" dictionaries learned by K-SVD or Sparse K-SVD, or than the predetermined overcomplete DCT dictionary. We recently developped a new sparse coding method adapted to this tree-structure to improve the results. Our dictionary learning method associated with this sparse coding method is also compared to other methods previously introduced in the recent litterature such as TSITD (Tree-Structured Iteration-Tuned Dictionary) or BITD (Basic Iteration-Tuned Dictionary) algorithms.

6.4. Distributed processing and robust communication

information theory, stochastic modelling, robust detection, maximum likelihood estimation, generalized likelihood ratio test, error and erasure resilient coding and decoding, multiple description coding, Slepian-Wolf coding, Wyner-Ziv coding, information theory, MAC channels

6.4.1. Loss concealment based on video inpainting

Participants: Mounira Ebdelli, Christine Guillemot, Ronan Le Boulch, Olivier Le Meur.

In 2011, we have started developing a loss concealment scheme based on a new video examplar-based inpainting algorithm. The developed video inpainting approach relies on a motion confidence-aided neighbor embedding techniques. Neighbor embedding approaches aim at approximating input vectors (or data points) as a linear combination of their neighbors. We have considered two neighbor embedding approaches namely locally linear embedding (LLE) and non-negative matrix factorization (NMF), in a way that each patch of the target region is inpainted with the best estimation provided using template matching, LLE and NMF. The motion confidence introduced in the neighbor embedding improves the robustness of the algorithm with limiting the error propagation effects which may result from uncertainties on the motion information of the unknown pixels to be estimated. Evaluations of the algorithm in a context of video editing (object removal) show natural looking videos with less annoying artifacts [24].

This approach has then been adapted to the context of loss concealment that is to estimate unknown pixels after decoding when the corresponding transport packets have been lost on the transmission network. For this purpose, a preprocessing step is proposed to estimate the motion information of each corrupted block using Bilinear Motion Field Interpolation (BMFI) before inpainting the texture. The BMFI method computes the missing motion vector of each pixel in the lost block as a weighted combination of motion vectors of neighboring blocks. The estimated motion information is also used to limit the search space for the best matching patches in a motion-compensated window. Experiments of the proposed approach on several videos show a PSNR average gain about 2dB compared to state-of-art methods [25]. The next step will be to assess the performance of the approach in a context of free moving camera videos. To deal with this problem, we propose to consider a panoramic image mosaics in order to estimate the background of the video before inpainting the missing part of the foreground objects.

6.4.2. Unequal Erasure Protection and Object Bundle Protection

Participant: Aline Roumy.

In 2011, we started a new collaboration on Unequal Erasure Protection (UEP) and Object Bundle Protection in the framework of the joint research lab Inria–Alcatel Lucent and the ANR ARSSO project. Protection is usually obtained by adding Forward error correction (FEC) to the object (or data) to be transmitted. However, when the object contains information with different importance levels (as in a video bitstream), providing a protection adapted to the importance of each subpart of the object, helps reducing the encoded bitrate. To implement UEP, traditional transport protocols based on FEC Schemes need to split the original object into say two sub-objects, one per important class, and to submit each sub-object separately to the FEC Scheme. This requires extra logic for splitting/gathering the data. A companion problem, is the case where the object size is smaller than the packetsize. In this case, FEC traditional approaches applied to each small object is wasting the bandwidth. An optimized solution consists in grouping the small objects with equal importance into a single file. This is the goal of object bundle protection. We proposed a novel method, called Generalized Object Encoding that can deal with both aspects [37], [38], [39]. In 2011, we analyzed our GOE approaches with average metrics such as average waiting time, average number of packets to be encoded. In 2012, we continued the analysis and considered memory requirements at the decoder [30].

6.4.3. Universal distributed coding

Participant: Aline Roumy.

In 2012, we started a new collaboration with Michel Kieffer and Elsa Dupraz (Supelec, L2S) on universal distributed source coding. Distributed source coding refers to the problem where several correlated sources need to be compressed without any cooperation at the encoders. Decoding is however performed jointly. This problem arises in sensor networks but also in video compression techniques, where the correlation between the successive frames is not directly used at the encoder as in [17], and are therefore seen as distributed. Traditional approaches (from an information theoretical but also practical point of view) assume that the correlation channel between the sources is perfectly known. Since this assumption is not satisfied in practice, a way to get around this is to use a feedback channel (from the decoder to the encoder), that can trigger the encoder. Instead, we consider universal distributed source coding, where the correlation channel is unknown and belongs to a class parametrized by some unknown parameter vector. In [23], we proposed four uncertainty models that depend on the partial knowledge we have on the correlation channel and derived the information theoretical bounds.

6.4.4. Super-resolution as a communication tool

Participants: Marco Bevilacqua, Christine Guillemot, Aline Roumy.

In 2012, we carried on the collaboration with Alcatel Lucent Bell Labs, represented by M-L. Alberi Morel, in the framework of a Joint Inria/Alcatel Lucent lab. In this work, we continued investigating super resolution (SR) as a potential tool to use in the context of video transmission. As SR refers to the task of producing a high-resolution (HR) image from one or several low-resolution (LR) input images, one can think of sending a LR video to adapt to the complexity constraint of the encoder and/or the bandwidth limitation of the network, and still being able to reconstruct a HR video at the encoder side, by applying a SR algorithm.

As a first step toward the more ambitious goal of compressing video through SR, we developed a novel method for single-image SR based on a neighbor embedding technique. In the neighbor embedding based SR procedure, the LR input image is first divided into small patches, namely sub-windows of image. Each input patch is approximated by a linear combination of its nearest neighbors (LR candidate patches) taken from a dictionary. Then, the corresponding HR output patch is created by combining similarly the corresponding HR candidates of the dictionary. The SR image is finally obtained by aggregating all the single HR patches reconstructed. A key point of this approach is represented by the above mentioned dictionary, which is a stored set of LR and HR patch correspondences extracted from training natural images.

The studies undertaken led us to have two publications in international conferences [20], [19]: ICASSP (International Conference on Acoustics, Speech, and Signal Processing) and BMVC (British Machine Vision Conference). In [20] we presented a neighbor embedding based SR method, by following the general scheme, but also introducing a new method to compute the weights of the linear combinations of patches. The weights of a certain input patch are computed as the result of a least squares problem with a nonnegative constraint. The so resulting nonnegative weights, that intuitevely represent a reasonable solution as they allow only additive combinations of patches, are shown to perform better than other weight computation methods described in the literature. The least squares problem is solved in a original fashion by means of SNMF, a tool for matrix factorization with one nonnegative factor. In [19] we refined the proposed algorithm, by focusing more on a low complexity target and by giving some theoretical insights about the choice of the nonnegative embedding. An analysis about the representation of the patches (either by the straight luminance values of its pixels or by some "features" conveniently computed) is also performed. The algorithm is shown to have better results, both in terms of quality performance and running time, than other similar SR algorithms that also adopt a one-pass procedure; and comparable visual results with respect to more sophisticated multi-pass algorithms, but still presenting a much reduced computational time. During the year, some other studies have been conducted, e.g. on the creation of the dictionary and on alternative ways to select the candidate patches from the dictionary. These extra studies, together with the already consolidated work of the published papers, represent the point of departure to the next step of designing a framework for video super resolution.

SMIS Project-Team

6. New Results

6.1. Embedded Data Management

Participants: Nicolas Anciaux, Luc Bouganim, Lionel Le Folgoc, Yanli Guo, Saliha Lallali, Philippe Pucheral, Iulian Sandu Popa, Shaoyi Yin.

Inspired by low cost economic models, this work draws the idea of a one-dollar database machine, with the objective to disseminate databases everywhere, up to the lightest smart objects. In contrast to traditional database machines relying on massively parallel architectures, the one-dollar database machine considers the cheapest form of computer available today: a microcontroller equipped with GBs size (external) Flash storage. Designing such a database machine is very challenging due to a combination of conflicting RAM and NAND Flash constraints. To tackle this challenge, this work proposes a new paradigm based on database serialization (managing all database structures in a pure sequential way) and stratification (restructuring them into strata when a scalability limit is reached).We show that a complete DBMS engine can be designed according to this paradigm and demonstrate the effectiveness of the approach through a performance evaluation. This work capitalizes on previous results related to the indexing of Flash resident data [16] and has also obvious connections with the more general study we are conducting on Flash-based data management (see Section 6.2). Partial elements of this solution have been demonstrated at [13]. In 2012, we have extended our previous results on indexation of flash resident data [IS] and we have proposed the design of a complete DBMS engine [DAPD] complying by nature with the conflicting RAM and NAND Flash constraints we are facing. Currently, we work at the extension of the embedded DBMS engine to support document data (e.g., text documents or any type of documents that are tagged)) and spatio-temporal data (e.g., vehicle trajectory data or any type of time-stamped and/or geo-located data).

6.2. Flash-based Data Management

Participants: Matias Bjørling, Philippe Bonnet, Luc Bouganim, Niv Dayan.

Solid State Drives (SSDs) are replacing magnetic disks as secondary storage for database management, as they offer orders of magnitude improvement in terms of bandwidth and latency. In terms of system design, the advent of SSDs raises considerable challenges. First, the storage chips, which are the basic component of a SSD, have widely different characteristics – e.g., copy-on-write, erase-before-write and page-addressability for flash chips vs. in-place update and byte-addressability for PCM chips. Second, SSDs are no longer a bottleneck in terms of I/O latency forcing streamlined execution throughout the I/O stack. Finally, SSDs provide a high degree of parallelism that must be leveraged to reach nominal bandwidth. This evolution puts database system researchers at a crossroad. The first option is to hang on to the current architecture where secondary storage is encapsulated behind a block device interface. This is the mainstream option both in industry and academia. This leaves the storage and OS communities with the responsibility to deal with the complexity introduced by SSDs in the hope that they will provide us with a robust, yet simple, performance model. We showed that this option amounts to building on quicksand. We illustrated our point by debunking some popular myths about flash devices and by pointing out mistakes in the papers we have published throughout the years. The second option is to abandon the simple abstraction of the block device interface and reconsider how database storage managers, operating system drivers and SSD controllers interact. We gave our vision of how modern database systems should interact with secondary storage. This approach requires a deep re-design of the database system architecture, which is the only viable option for database system researchers to avoid becoming irrelevant. This work started at the end of 2011 and was published at CIDR'13 [20], in cooperation with the IT University of Copenhagen.

6.3. Minimal Exposure

Participants: Nicolas Anciaux, Walid Bezza, Danae Boutara, Benjamin Nguyen, Michalis Vazirgiannis.

When users request a service, the service provider usually asks for personal documents to tailor its service to the specific situation of the applicant. For example, the rate and duration of consumer's loans are usually adapted depending on the risk based on the income, assets or past lines of credits of the borrower. In practice, an excessive amount of personal data is collected and stored. Indeed, a paradox is at the root of this problem: service providers require users to expose data in order to determine whether that data is needed or not to achieve the purpose of the service. We explore a reverse approach, where service providers would publicly describe the data they require to complete their task, and where software (placed, depending on the context, on the client, on the server, or in a trusted hardware component) would use those descriptions to determine a minimum subset of information to expose. In 2012, we have presented our general framework called Minimum Exposure [14], we have modelled the underlying problem (for simple tasks) and proposed resolution algorithms [19], [24], and we have addressed the case of multi-label classifiers [18]. In the short term, we plan to adapt the minimum exposure architecture to support hidden decision rules using smart cards. Then, we will investigate new privacy metrics to capture the degree of exposure of sets of personal data items better.

6.4. Secure Global Computing on Asymmetric Architecture

Participants: Tristan Allard, Benjamin Nguyen, Philippe Pucheral, Quoc-Cuong To.

This research direction is based on the asymmetric architecture, composed of a powerful, available and untrusted computing infrastructure (server or cloud), and a large set of low powered, highly disconnected trusted devices. Trust is assumed ad hoc and can be justified by the use of secure tokens, open source software, friend relationships etc. In our work, we use tamper resistant secure tokens running trusted software, which provide a high degree of trust, due to the overwhelming cost of hardware tampering. The main difficulty on such an architecture is global processing i.e. constructing aggregate data from the individual records, because the entity in charge of executing the global computation is untrusted. Given our large scale data centric applications (e.g. nationwide surveys), we also discard solutions based on secure multi-party computation, which do not scale. We have studied the execution of Privacy Preserving Data Publishing (PPDP) algorithms on such an architecture, and provided generic protocols to deal with all kinds of PPDP algorithms, which are robust against honest-but-curious and malicious adversaries. This work is an extension of [31] We are now studying more generally the execution of SQL "Group by" queries on this architecture, which is the topic of Quoc-Cuong To's Ph.D. thesis started in sept. 2012. We have published preliminary results on this novel problem in [23], which adapts the techniques proposed in [31].

6.5. Trusted Cell Data Management

Participants: Nicolas Anciaux, Philippe Bonnet, Luc Bouganim, Benjamin Nguyen, Philippe Pucheral, Iulian Sandu Popa.

With the convergence of mobile communications, sensors and online social networks technologies, we are witnessing an exponential increase in the creation and consumption of personal data. Such data is volunteered by users, automatically captured by sensors or inferred from existing data. Today, there is a wide consensus that individuals should have increased control on how their personal data is collected, managed and shared. Yet there is no appropriate technical solution to implement such personal data services: centralized solutions sacrifice security for innovative applications, while decentralized solutions sacrifice innovative applications for security. In this work, we argue that the advent of secure hardware in all personal IT devices, at the edges of the Internet, could trigger a sea change. We propose the vision of trusted cells: personal data servers running on secure smart phones, set-top boxes, secure portable tokens or smart cards to form a global, decentralized data platform that provides security yet enables innovative applications. We motivate our approach, describe the trusted cells architecture and define a range of challenges for future research in a paper published at CIDR'13 (Int. Conf on Innovative Data Systems Research) [17].

6.6. Experiment in the medical field

Participants: Nicolas Anciaux, Luc Bouganim, Philippe Pucheral, Alexei Troussov.

The PlugDB engine is being experimented in the field since September 2011 to implement a secure and portable medical-social folder. The objective is to improve the coordination of medical care and social services provided at home for dependent people. Details related to this experiment are given in Section 7.2 . While this action did not generate new academic results (though it helped us validating some previous results), it imposed us a strong investment in terms of test and optimization for our prototype and in terms of communication to promote this experiment at the regional level.

STARS Team

6. New Results

6.1. Introduction

This year Stars has proposed new algorithms related to its three main research axes : perception for activity recognition, semantic activity recognition and software engineering for activity recognition.

6.1.1. Perception for Activity Recognition

Participants: Julien Badie, Slawomir Bak, Vasanth Bathrinarayanan, Piotr Bilinski, Bernard Boulay, François Brémond, Sorana Capalnean, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Eben Freeman, Carolina Garate, Jihed Joober, Vaibhav Katiyar, Ratnesh Kumar, Srinidhi Mukanahallipatna, Sabine Moisan, Silviu Serban, Malik Souded, Anh Tuan Nghiem, Monique Thonnat, Sofia Zaidenberg.

This year Stars has extended an efficient algorithm for detecting people. We have also proposed a new algorithm for re-identification of people through a camera network. We have realized a new algorithm for the recognition of short actions and validated also its performance on several benchmarking databases (e.g. ADL). We have improved a generic event recognition algorithm by handling event uncertainty at several processing levels. More precisely, the new results for perception for activity recognition concern:

- Image Compression and Modelization (6.2)
- Background Subtraction (6.3)
- Fiber Based Video Segmentation (6.4)
- Enforcement of Monotonous Shape Growth/Shrinkage in Video Segmentation (6.5)
- Dynamic and Robust Object Tracking in a Single Camera View (6.6)
- Optimized Cascade of Classifiers for People Detection Using Covariance Features (6.7)
- Learning to Match Appearances by Correlations in a Covariance Metric Space (6.8)
- Recovering Tracking Errors with Human Re-identification (6.9)
- Human Action Recognition in Videos (6.10)
- Group Interaction and Group Tracking for Video-surveillance in Underground Railway Stations (6.11)
- Crowd Event Monitoring Using Texture and Motion Analysis (6.12)
- Detecting Falling People (6.13)
- People Detection Framework (6.14)

6.1.2. Semantic Activity Recognition

Participants: Sorana Capalnean, Guillaume Charpiat, Cintia Corti, Carlos -Fernando Crispim Junior, Hervé Falciani, Baptiste Fosty, Qioa Ma, Firat Ozemir, Jose-Luis Patino Vilchis, Guido-Tomas Pusiol, Rim Romdhame, Bertrand Simon, Abhineshwar Tomar.

Concerning semantic activity recognition, the contributions are :

- A Model-based Framework for Activity Recognition of Older People using Multiple sensors (6.15)
- Activity Recognition for Older People using Kinect (6.16)
- Descriptors of Depth-Camera Videos for Alzheimer Symptom Detection (6.17)
- Online Activity Learning from Subway Surveillance Videos (6.18)
- Automatic Activity Detection Modeling and Recognition: ADMR (6.19)

6.1.3. Software Engineering for Activity Recognition

Participants: François Brémond, Daniel Gaffé, Julien Gueytat, Baptiste Fosty, Sabine Moisan, Anh tuan Nghiem, Annie Ressouche, Jean-Paul Rigault, Leonardo Rocha, Luis-Emiliano Sanchez, Swaminathan Sankaranarayanan.

This year Stars has continued the development of the SUP platform. This latter is the backbone of the team experiments to implement the new algorithms. We continue to improve our meta-modelling approach to support the development of video surveillance applications based on SUP. This year we have focused on an architecture for run time adaptation and on metrics to drive dynamic architecture changes. We continue the development of a scenario analysis module (SAM) relying on formal methods to support activity recognition in SUP platform. We improve the theoretical foundations of CLEM toolkit and we rely on it to build SAM. Finally, we are improving the way we perform adaptation in the definition of a multiple services for device adaptive platform for scenario recognition.

The contributions for this research axis are:

- SUP Software Platform (6.20)
- Qualitative Evaluation of Detection and Tracking Performance (6.21)
- Model-Driven Engineering and Video-surveillance (6.22)
- Synchronous Modelling and Activity Recognition (6.23)

6.2. Image Compression and Modelization

Participants: Guillaume Charpiat, Eben Freeman.

Recent results in statistical learning have established the best strategy to combine several advices from different experts, for the problem of sequential prediction of times series. The notions of prediction and compression are tightly linked, in that a good predictor can be turned into a good compressor via entropy coding (such as Huffman coding or arithmetic coding), based on the predicted probabilities of the events to come : the more predictable an event E is, the easier to compress it will be, with coding cost $-\log(p(E))$ with such techniques.

The initial idea here, by Yann Ollivier (TAO team), within a collaboration with G. Charpiat and Jamal Atif (TAO team), was to adapt these results to the case of image compression, where time series are replaced with 2D series of pixel colors, and where experts are predictors of the color of a pixel given the colors of neighbors. The main difference is that there is no canonical physically-relevant 1D ordering of the pixels in an image, so that a sequential order (of the pixels to predict their colors) had to be defined first. Preliminary results with a hierarchical ordering scheme already competed with standard techniques in lossless compression (png, lossless jpeg2000).

During his internship in the Stars team, Eben Freeman developed this approach, by building relevant experts able to predict a variety of image features (regions of homogeneous color, edges, noise, ...). We also considered random orderings of pixels, using kernels to express probabilities in a spatially-coherent manner. Using such modellings of images with experts, we were also able to generate new images, that are typical of these models, and show more structure than the ones associated to standard compression schemes (typical images highly compressed).

6.3. Background Subtraction

Participants: Vasanth Bathrinarayanan, Anh-Tuan Nghiem, Duc-Phu CHAU, François Brémond.

Keywords: Gaussian Mixture Model, Shadow removal, Parameter controller, Codebook model, Context based information

6.3.1. Statistical Background Subtraction for Video Surveillance Platform

Anh-Tuan Nghiem work on background subtraction is an extended version of Gaussian Mixture Models [73]. The algorithm compares each pixel of current frame to background representation which is developed based on the pixel information from previous frames. It includes shadow and highlight removal to give better results. Selective background updating method based on the feedback from the object detection helps to better model background and remove noise and ghosts.

Figure 10 shows a sample illustration of the output of the background subtraction, where blue are foreground pixels and red are shadow or illumination change pixels and a green bounding box is a foreground blob. Also we have compared our algorithm with few other such as OpenCV and also IDIAP's background subtraction(not tuned perfectly, used default parameters) and the results are shown in Figure 11 where the green background refers to best performance of the comparisons. This evaluation is done on PETS 2009 data-set with our obtained foreground blobs to the manually annotated bounding boxes of people.

6.3.2. Parameter controller using Contextual features

The above method has some parameters that has to be tuned every time for each video, which is a time consuming work. The work of Chau et al [59] learns the contextual information from the video and controls object tracking algorithm parameters during the run-time of the algorithm. This approach is at preliminary stage for background subtraction algorithm to automatically adapt parameters. These parameters are learned as described in the offline learning process block diagram 12 over several ground truth videos and clustered into a database. The contextual feature which are used presently include object density, occlusion, contrast, 2D area, contrast variance, 2D area variance. Figure 13 shows a sample of video chunks based on contextual feature similarity for a video from caviar data-set.

The controller's preliminary results are promising and we are experimenting and evaluating with different features to learn the parameters. The results will be published in upcoming top computer vision conferences.

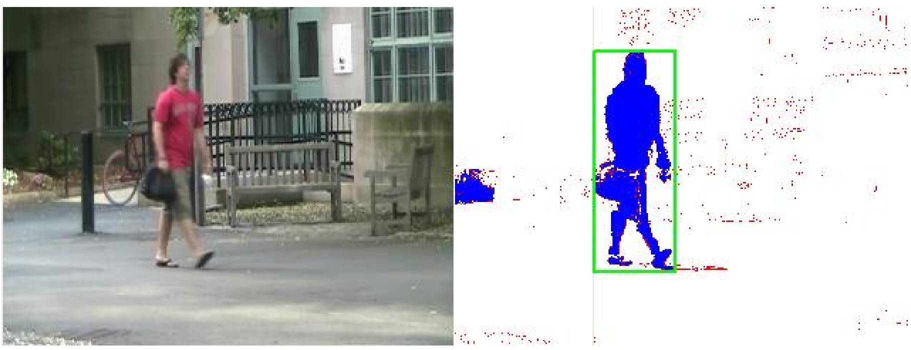


Figure 10. Sample illustration of output of EGMM

6.4. Fiber Based Video Segmentation

Participants: Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

Keywords: Video Volume, Fibers, Trajectory

PETS dataset Evaluation on Background subtraction blob with GT of people--			
Metrics	**IDIAP**	**SUP**	**OPENCV**
Metric M1.1			
Global results:			
Number of True Positives	3761	4046	4057
Number of False Positives	1807	317	1099
Number of False Negatives 0	1070	793	774
Precision (mean by frame)	0.71	0.93	0.81
Sensitivity 0 (mean by frame)	0.79	0.84	0.85
Precision (global)	0.68	0.93	0.79
Sensitivity 0 (global)	0.78	0.84	0.84

Figure 11. Evaluation with some background subtraction algorithms

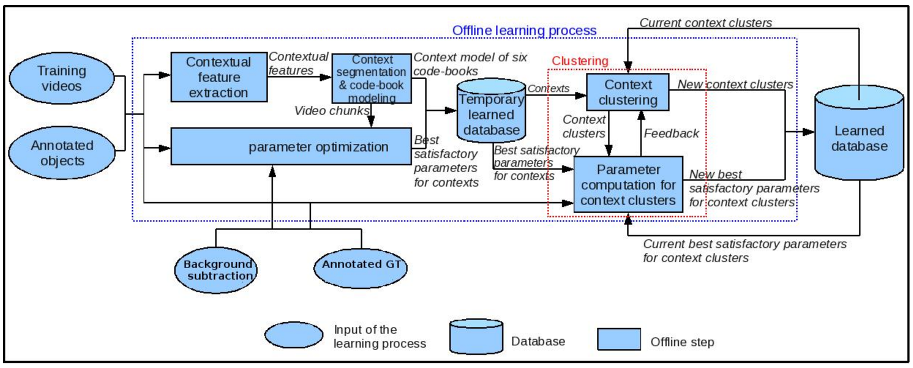


Figure 12. Block representation of the offline learning scheme to build a database

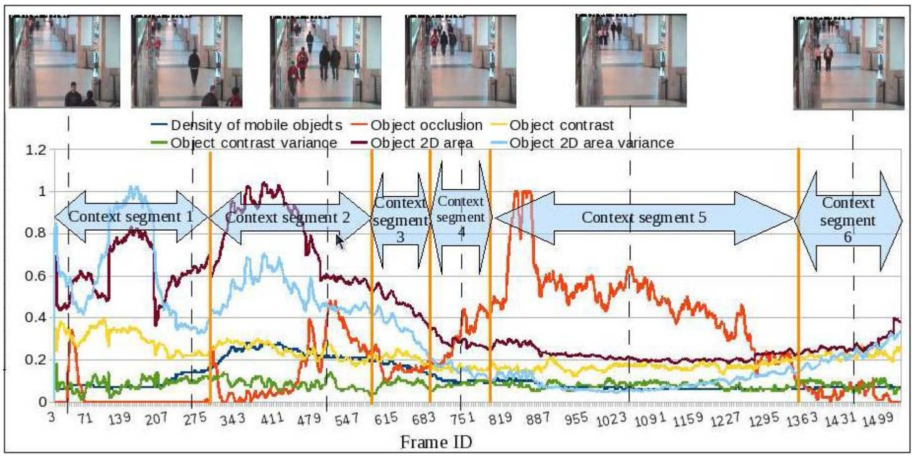


Figure 13. Context segmentation of the sequence ThreePastShop2cor (belonging to the Caviar dataset). The context segments are separated by the vertical orange lines. The control parameters are then learned for each context segment

The aim of this work is to segment objects in videos by considering videos as 3D volumetric data (2D×time). Figure 14 shows an input video and its corresponding partition in terms of fiber at a particular hierarchy level. Particularly, it shows 2D slices of a video volume. Bottom right corner of each figure shows the current temporal depth in the volume, while top right shows the X-time slice and bottom left shows Y-time slice. In this 3D representation of videos, points of static background form straight lines of homogeneous intensity over time, while points of moving objects form curved lines. Analogically to the fibers in MRI images of human brains, we term *fibers*, these straight and curved lines of homogeneous intensity. So, in our case, to segment the whole video volume data, we are interested in a dense estimation of fibers involving all pixels.

Initial fibers are built using correspondences computing algorithms like optical flow and descriptor matching. As these algorithms are reliable near corners and edges, we build fibers at these locations for a video. Our subsequent goal is to partition this video in terms of fibers built, by extending them (both spatially and temporally) to the rest of the video.
To extend fibers, we compute geodesics from pixels (not belonging to the initially built fibers) to fibers. For a reliable extension, the cost of moving along a geodesic is proportional to the trajectory similarity of a pixel *wrt* a fiber, wherein a pixel trajectory is similar to the fiber trajectory. This cost function quantifies the color homogeneity of a pixel trajectory along with its color similarity *wrt* a fiber. A pixel is then associated to a fiber for which this cost is minimum.
With the above mentioned steps we obtain a partition of a video in terms of fibers wherein we have a trajectory associated with each pixel. This hierarchical partition provides a mid-level representation of a video, which can be seen as a facilitator or a pre-processing step towards higher level video understanding systems *eg* activity recognition.

6.5. Enforcement of Monotonous Shape Growth/Shrinkage in Video Segmentation

Participant: Guillaume Charpiat.

keywords: graph cuts, video segmentation, shape growth

Figure 14. ***Left:*** *Input Video and Spatio-Temporal Slices.* ***Right:****Segmented Results at a Particular Hierarchy Level*

The segmentation of noisy videos or time series is a difficult problem, not to say an impossible or ill-posed task when the noise level is very high. While individual frames can be analysed independently, time coherence in image sequences provides a lot of information not available for a single image. Most of the state-of-art works explored short-term temporal continuity for object segmentation in image sequences, *i.e.*, each next frame is segmented by using information from one or several images at previous time points. It is, however, more advantageous to simultaneously segment many frames in the data set, so that segmentation of the entire image set supports each of the individual segmentations.

In this work, we focus on segmenting shapes in image sequences which only grow or shrink in time, and on making use of this knowledge as a constraint to help the segmentation process. Examples of growing shapes are forest fires in satellite images and organ development in medical imaging. We propose a segmentation framework based on graph cuts for the joint segmentation of a multi-dimensional image set. By minimizing an energy computed on the resulting spatio-temporal graph of the image sequence, the proposed method yields a *globally optimal solution*, and runs in practice in linear complexity in the total number of pixels.

Two applications are performed. First, with Yuliya Tarabalka (Ayin team), we segment multiyear sea ice floes in a set of satellite images acquired through different satellite sensors, after rigid alignment (see Figure 15). The method returns accurate melting profiles of sea ice, which is important for building climate models. The second application, with Bjoern Menze (ETH Zurich, also MIT and collaborator of Asclepios team), deals with the segmentation of brain tumors from longitudinal sets of multimodal MRI volumes. In this task we impose an additional inter-modal inclusion constraint for joint segmentation of different image sequences, finally also returning highly sensitive time-volume plots of tumor growth.

6.6. Dynamic and Robust Object Tracking in a Single Camera View

Participants: Duc-Phu Chau, Julien Badie, François Brémond, Monique Thonnat.

Keywords: Object tracking, online parameter tuning, controller, self-adaptation and machine learning

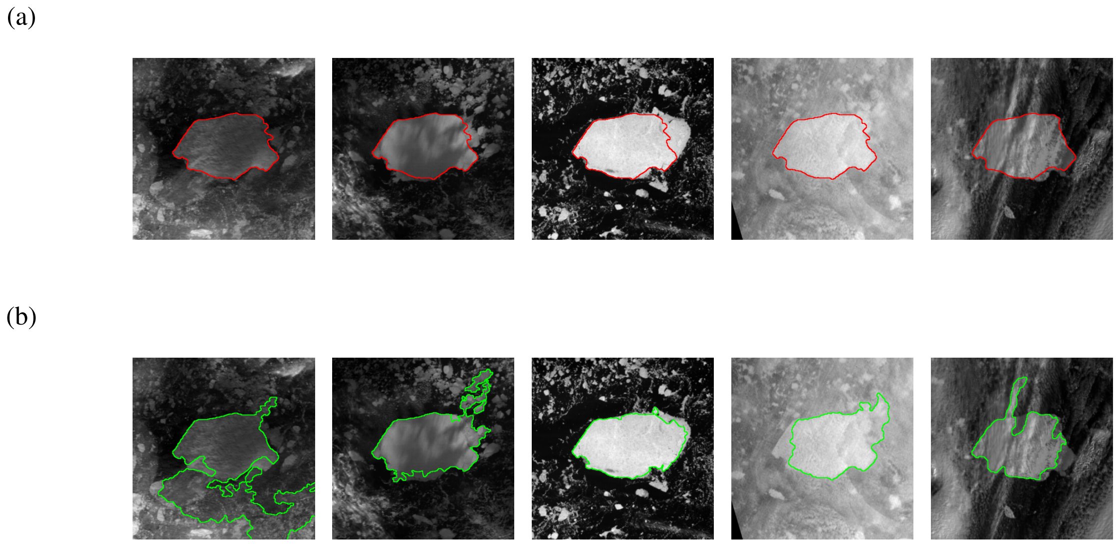


Figure 15. (a) Aligned satellite images captured each four days superposed with segmentation contours computed by our approach. (b) Segmentation contours for images (a) obtained by applying graph cut segmentation to each image at a single time moment. Note that the segmentations (a) are pixelwise precise, and that the white regions surrounding sometimes the boundaries are other ice blocks, agglomerating temporarily only, thus correctly labelled. Hence the importance of enforcing time coherence.

Object tracking quality usually depends on video scene conditions (e.g. illumination, density of objects, object occlusion level). In order to overcome this limitation, we present a new control approach to adapt the object tracking process to the scene condition variations. The proposed approach is composed of two tasks.

The objective of the first task is to select a convenient tracker for each mobile object among a Kanade-Lucas-Tomasi-based (KLT) tracker and a discriminative appearance-based tracker. The KLT feature tracker is used to decide whether an object is correctly detected. For badly detected objects, the KLT feature tracking is performed to correct object detection. A decision task is then performed using a Dynamic Bayesian Network (DBN) to select the best tracker among the discriminative appearance and KLT trackers.

The objective of the second task is to tune online the tracker parameters to cope with the tracking context variations. The tracking context, or context, of a video sequence is defined as a set of six features: density of mobile objects, their occlusion level, their contrast with regard to the surrounding background, their contrast variance, their 2D area and their 2D area variance. Each contextual feature is represented by a code-book model. In an offline phase, training video sequences are classified by clustering their contextual features. Each context cluster is then associated with satisfactory tracking parameters. In the online control phase, once a context change is detected, the tracking parameters are tuned using the learned values. This work has been published in [29], [35].

We have tested the proposed approach on several public datasets such as Caviar and PETS. Figure 16 illustrates the results of the object detection correction using the KLT feature tracker.

Figure 17 illustrates the tracking output for a Caviar video (on the left image) and for a PETS video (on the right image). The experimental results show that our method gets the best performance compared to some recent state of the art trackers.

Table 1 presents the tracking results for 20 videos from the Caviar dataset. The proposed approach obtains the best MT value (i.e. mostly tracked trajectories) compared to some recent state of the art trackers.

Figure 16. Illustration of the object detection correction for a Caviar video. The green bounding box is the output of the object detection process. The red bounding boxes are the results of the detection correction task.

a. b.

Figure 17. Tracking results for Caviar and PETS videos

Table 1. Tracking results on the Caviar dataset. MT: Mostly tracked trajectories, higher is better. PT: Partially tracked trajectories. ML: Most lost trajectories, lower is better. The best values are printed bold.

Method	MT (%)	PT (%)	ML (%)
Zhang et al., CVPR 2008 [89]	85.7	10.7	3.6
Li et al., CVPR 2009 [71]	84.6	14.0	1.4
Kuo et al., CVPR 2010 [69]	84.6	14.7	**0.7**
Proposed approach	**86.4**	10.6	3.0

Table 2 presents the tracking results of the proposed approach and three recent approaches [56], [82], [67] for a PETS video. With the proposed approach, we obtain the best values in both metrics MOTA (i.e. Multi-object tracking accuracy) and MOTP (i.e. Multi-object tracking precision). The authors in [56], [82], [67] do not present the tracking results with the MT, PT and ML metrics.

Table 2. Tracking results on the PETS sequence S2.L1, camera view 1, sequence time 12.34. MOTA: Multi-object tracking accuracy, higher is better. MOTP: Multi-object tracking precision, higher is better. The best values are printed bold.

Method	MOTA	MOTP	MT (%)	PT (%)	ML (%)
Berclaz et al., PAMI 2011 [56]	0.80	0.58	-	-	-
Shitrit et al., ICCV 2011 [82]	0.81	0.58	-	-	-
Henriques et al., ICCV 2011 [67]	0.85	0.69	-	-	-
Proposed approach	**0.86**	**0.72**	71.43	19.05	9.52

6.7. Optimized Cascade of Classifiers for People Detection Using Covariance Features

Participants: Malik Souded, François Brémond.

keywords: People detection, Covariance descriptor, LogitBoost.

We propose a new method to optimize a state of the art approach for people detection, which is based on classification on Riemannian manifolds using covariance matrices in a boosting scheme. Our approach makes training and detection faster while maintaining equivalent or better performances. This optimisation is achieved by clustering negative samples before training, providing a smaller number of cascade levels and less weak classifiers in most levels in comparison with the original approach.

Our approach is based on Tuzel et al. [86] work which was improved by Yao et al. [87]. We keep the same scheme to achieve our people detector: train a cascade of classifiers based on covariance descriptors, using a Logitboost training algorithm which was modified by Tuzel et al. to deal with the Riemannian manifolds metrics and using the operators which were presented in [75]. In fact, Covariance matrices do not belong to vector space but to the Riemannian manifold of (d x d) symmetric positive definite matrices. The trained cascade of classifiers is applied for detection after training.

We propose an additional step to speed up training and detection process. We propose to apply a clustering step on negative training dataset before training the classifiers. This clustering step is performed both in Riemannian manifold and in the vector space of mapped covariance matrices, using the operators and metrics previously cited.

The idea consists in regrouping all similar negative samples, with regard to their covariance information, into decreasing size clusters. Each classifier of the cascade is trained on one cluster, specializing this classifier for a given kind of covariance information, and then, speeding up the training step and providing shorter classifier, which accelerate its response when applied on image. In the same time, the specialization of each cascade classifier shortens the cascade too, speeding up the detection (see Figure 18 and Figure 19).

A paper describing this approach has been accepted in VISAPP 2013 conference [50].

6.8. Learning to Match Appearances by Correlations in a Covariance Metric Space

Participants: Sławomir Bąk, Guillaume Charpiat, Etienne Corvée, Francois Brémond, Monique Thonnat.

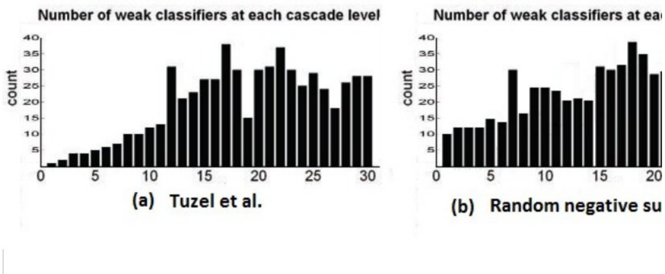


Figure 18. Comparison of classifiers cascade structure

(a) (b) (c)

Figure 19. Cascade classifiers: some detection results

keywords: covariance matrix, re-identification, appearance matching

This work addresses the problem of appearance matching across disjoint camera views. Significant appearance changes, caused by variations in view angle, illumination and object pose, make the problem challenging.

We propose to formulate the appearance matching problem as the task of learning a model that selects the most descriptive features for a specific class of objects. Our main idea is that different regions of the object appearance ought to be matched using various strategies to obtain a distinctive representation. Extracting region-dependent features allows us to characterize the appearance of a given object class (*e.g.* class of humans) in a more efficient and informative way. Different kinds of features characterizing various regions of an object is fundamental to our appearance matching method.

We propose to model the object appearance using covariance descriptor yielding rotation and illumination invariance. Covariance descriptor has already been successfully used in the literature for appearance matching. In contrast to state of the art approaches, we do not define *a priori* feature vector for extracting covariance, but we learn which features are the most descriptive and distinctive depending on their localization in the object appearance (see figure 20). Learning is performed in a covariance metric space using an entropy-driven criterion. Characterizing a specific class of objects, we select only essential features for this class, removing irrelevant redundancy from covariance feature vectors and ensuring low computational cost.

The proposed technique has been successfully applied to the person re-identification problem, in which a human appearance has to be matched across non-overlapping cameras [34]. We demonstrated that: (1) by using different kinds of covariance features *w.r.t.* the region of an object, we obtain clear improvement in appearance matching performance; (2) our method outperforms state of the art methods in the context of pedestrian recognition on publicly available datasets (i-LIDS-119, i-LIDS-MA and i-LIDS-AA); (3) using 4×4 covariance matrices we significantly speed-up the processing time offering an efficient and distinctive representation of the object appearance.

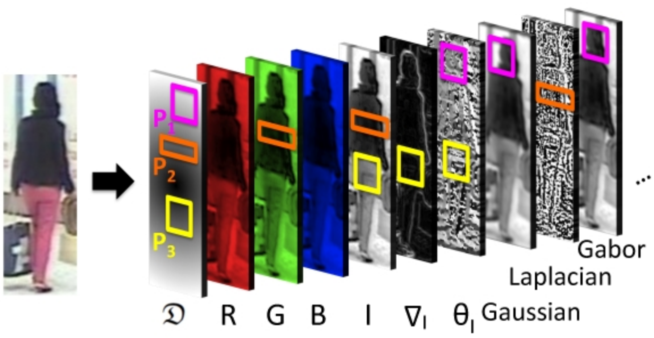


Figure 20. Example of three different covariance features. Every covariance is extracted from a region (P), distance layer (D) and three channel functions (e.g. bottom covariance feature is extracted from region P_3 using layers: D, I-intensity, ∇_I-gradient magnitude and θ_I-gradient orientation).

6.9. Recovering Tracking Errors with Human Re-identification

Participants: Julien Badie, Slawomir Bak, Duc-Phu Chau, François Brémond, Monique Thonnat.

keywords: tracking error correction, re-identification

This work addresses the problem of people tracking at long range even if the target people are lost several times by the tracking algorithm. We have identified two main reasons for tracking interruption. The first one concerns interruptions that can be quickly recovered, which includes short mis-detections, occlusions with other persons or static obstacles. The second one occurs when a person is occluded or mis-detected for a long time or when the person leaves the scene and comes back latter. Our main objective is to design a framework that can track people even if their trajectory is very segmented and/or associated with different IDs. We called this problem the global tracking challenge (see Figure 21).

Figure 21. The global tracking challenge : correcting errors due to occlusions (ID 142 on the first frame becomes 147 on the last frame) and tracking people that are leaving the scene and reentering (ID 133 on the first frame becomes 151 on the last frame).

In order to describe a person's tracklet (segment of trajectory), we use a visual signature called Mean Riemannian Covariance Grid and a discriminative method to emphasize the main differences between each tracklet. This step improves the reliability and the accuracy of the results. By computing the distance between the visual signatures, we are able to link tracklets belonging to the same person into a tracklet cluster. Only tuples of tracklets that are not overlapping each other are used as initial candidates. Then, we use Mean Shift to create the clusters.

We evaluated this method on several datasets (i-LIDS, Caviar, PETS 2009). We have shown that our approach can perform as well as the other state of the art methods on Caviar and can perform better on i-LIDS. On PETS 2009 dataset, our approach performs better than standard tracker but cannot be compared with the best state of the art methods due to unadapted metrics.

This approach is described in detail in two articles : one published in ICIP 2012 [35], which is focused on computing the covariance signature and the way to discriminate it and the other one published in PETS 2012 workshop (part of AVSS 2012 conference) [33], which is focused on the method to link the tracklets.

This work will be added to a more general tracking controller that should be able to detect several kinds of detection and tracking errors and try to correct them.

6.10. Human Action Recognition in Videos

Participants: Piotr Bilinski, François Brémond.

keywords: Action Recognition, Contextual Features, Pairwise Features, Relative Tracklets, Spatio-Temporal Interest Points, Tracklets, Head Estimation.

The goal of this work is to automatically recognize human actions and activities in diverse and realistic video settings.

Over the last few years, the bag-of-words approach has become a popular method to represent video actions. However, it only represents a global distribution of features and thus might not be discriminative enough. In particular, the bag-of-words model does not use information about: local density of features, pairwise relations among the features, relative position of features and space-time order of features. Therefore, we propose three new, higher-level feature representations that are based on commonly extracted features (e.g. spatio-temporal interest points used to evaluate the first two feature representations or tracklets used to evaluate the last approach). Our representations are designed to capture information not taken into account by the model, and thus to overcome its limitations.

In the first method, we propose new and complex contextual features that encode spatio-temporal distribution of commonly extracted features. Our feature representation captures not only global statistics of features but also local density of features, pairwise relations among the features and space-time order of local features. Using two benchmark datasets for human action recognition, we demonstrate that our representation enhances the discriminative power of commonly extracted features and improves action recognition performance, achieving 96.16% recognition rate on popular KTH action dataset and 93.33% on challenging ADL dataset. This work has been published in [36].

In the second approach, we design new representation of features encoding statistics of pairwise co-occurring local spatio-temporal features. This representation focuses on pairwise relations among the features. In particular, we introduce the geometric information to the model and associate geometric relations among the features with appearance relations among the features. Despite that local density of features and space-time order of local features are not captured, we are able to achieve similar results on the KTH dataset (96.30% recognition rate) and 82.05% recognition rate on UCF-ARG dataset. An additional advantage of this method is to reduce the processing time of training the model from one week on a PC cluster to one day. This work has been published in [37].

In the third approach, we propose a new feature representation based on point tracklets and a new head estimation algorithm. Our representation captures a global distribution of tracklets and relative positions of tracklet points according to the estimated head position. Our approach has been evaluated on three datasets, including KTH, ADL, and our locally collected Hospital dataset. This new dataset has been created in cooperation with the CHU Nice Hospital. It contains people performing daily living activities such as: standing up, sitting down, walking, reading a magazine, *etc*. Sample frames with extracted tracklets from video sequences of the ADL and Hospital datasets are illustrated on Figure 22 . Consistently, experiments show that our representation enhances the discriminative power of tracklet features and improves action recognition performance. This work has been accepted for publication in [38].

Figure 22. Sample frames with extracted tracklets from video sequences of the ADL (left column) and Hospital (right column) datasets.

6.11. Group Interaction and Group Tracking for Video-surveillance in Underground Railway Stations

Participants: Sofia Zaidenberg, Bernard Boulay, Carolina Garate, Duc-Phu Chau, Etienne Corvée, François Brémond.

Keywords: events detection, behaviour recognition,automatic video understanding, tracking

One goal in the European project VANAHEIM is the tracking of groups of people. Based on frame to frame mobile object tracking, we try to detect which mobiles form a group and to follow the group through its lifetime. We define a group of people as two or more people being close to each other and having similar trajectories (speed and direction). The dynamics of a group can be more or less erratic: people may join or split from the group, one or more can disappear temporarily (occlusion or disappearance from the field of view) but reappear and still be part of the group. The motion detector which detects and labels mobile objects may also fail (misdetections or wrong labels). Analysing trajectories over a temporal window allows handling this instability more robustly. We use the event-description language described in [88] to define events, described using basic group properties such as size, type of trajectory or number and density of people and perform the recognition of events and behaviours such as violence or vandalism (alarming events) or a queue at the vending machine (non-alarming events).

The group tracking approach uses Mean-Shift clustering of trajectories to create groups. Two or more individuals are associated in a group if their trajectories have been clustered together by the Mean-Shift algorithm. The trajectories are given by the long-term tracker described in [60]. Each trajectory is composed of a person's positions (x, y) on the ground plane (in 3D) over the time window, and of their speed at each frame in the time window. Positions and speed are normalized using the minimum and maximum possible values (0 and $10m/s$ for the speed and the field of view of the camera for the position). The Mean-Shift algorithm requires a *tolerance* parameter which is set to 0.1, meaning that trajectories need to be distant by less than 10% of the maximum to be grouped.

Figure 23. Example of a group composed of non-similar individual trajectories.

As shown in Figure 23 , people in a group might not always have similar trajectories. For this reason, a group is also created when people are very close. A group is described by its coherence, a value calculated from the average distances of group members, their speed similarity and direction similarity. The update phase of the group uses the coherence value. A member will be kept in a group as long as the group coherence is above a threshold. This way, a member can temporarily move apart (for instance to buy a ticket at the vending machine) without being separated from the group.

This work has been applied to the benchmark CAVIAR dataset for testing, using the provided ground truth for evaluation. This dataset is composed of two parts: acted scenes in the Inria hall (9 sequences of 665 frames in average) and not acted recordings from a shopping mall corridor (7 sequences processed of 1722 frames in average). The following scenarios have been defined using the event-description language of [88]: *fighting*, *split up*, *joining*, *shop enter*, *shop exit*, *browsing*. These scenarios have been recognized in the videos with a high success rate (94%). The results of this evaluation and the above described method have been published in [45].

The group tracking algorithm is integrated at both Torino and Paris testing sites and runs in real time on live video streams. The global VANAHEIM system has been presented as a demonstration at the ECCV 2012 conference. A demonstration video has been compiled from the results of the group tracking on 60 sequences from the Paris subway showing interesting groups with various activities such as *waiting*, *walking*, *lost*, *kids* and *lively*.

6.12. Crowd Event Monitoring Using Texture and Motion Analysis

Participants: Vaibhav Katiyar, Jihed Joober, François Brémond.

keywords: Crowd Event, Texture Analysis, GLCM, Optical Flow

The aim of this work is to monitor crowd event using crowd density, change of speed and orientation of group of people. For reducing complexity we are using human density rather than individual human detection and tracking. In this study Human density is quantified mainly into three groups- (1) Empty (2) Sparse (3) Dense. These are approximated by calculating Haralick features from Grey Level Co-occurrence Matrix (GLCM).

We use Optical flow for getting motion information like current speed and orientation of selected FAST feature points. Subsequently we used this information for classifying crowd behaviour into normal or abnormal categories wherein we seek for sudden change in speed or orientation heterogeneity for abnormal behaviour.

In future work this abnormal behaviour may further be classified into different events like Running, Collecting, Dispersion, Stopping/Blocking.

6.13. Detecting Falling People

Participants: Etienne Corvee, Francois Bremond.

keywords: fall, tracking, event

We have developed a people falling algorithm based on our object detection and tracking algorithm [58] and using our ontology based event detector [57]. These algorithms extract moving object trajectories from videos and triggers alarms whenever the people activity fits event models. Most surveillance systems use a multi Gaussian technique [83] to model background scene pixels. This technique is very efficient in detecting in real-time moving objects in scenes captured by a static camera, with low level of shadows, few persons interacting in the scene and with as few as possible illumination changes. This technique does not analyse the content of the moving pixels but simply assign them as foreground or background pixels.

Many state of the art algorithms exist that can recognize objects such as a person human shape, a head, a face or a couch. However, these algorithms are quite time consuming or the database used for training the system is not well adapted to our application domain. For example, people detection algorithms use databases containing thousands of image instances of standing or walking persons taken by camera from a certain distance from the persons and from a facing position. In our indoor monitoring application, cameras are located on the roof with high tilt angle so that most of the scene (e.g.rooms) is viewed. With such camera spatial configuration, the image of a person on the screen rarely corresponds to the person images in the training database. In addition, people are often occluded by the image border (the image of the full body is not available), image distortion needs to be corrected and people often have poses that are not present in the database (e.g. a person bending or sitting).

Using our multi Gaussian technique [74], after having calibrated a camera scene, a detected object is associated with a 3D width and height in two positions : the standing and lying positions. This 3D information is checked against 3D human model and any object is then labelled as either a standing person, a lying person or unknown. Many 3D filtering thresholds are used ; for example, object speed should not be greater than a human possible running speed. Second, we use an ontology based event detector to build a hierarchy of event model complexity. We detect people to have fallen on the floor if the object has been detected as a person on the floor and outside the bed and couch for at least several seconds consecutively. An example of a fallen person is shown in Figure 24 .

Figure 24. Detection of a fallen person.

6.14. People Detection Framework

Participants: Srinidhi Mukanahallipatna, Silviu-Tudor Serban, François Brémond.

keywords: LBP, Adaboost, Cascades

We present a new framework called COFROD (Comprehensive Optimization Framework for Real-time Object Detection) for object detection that focuses on improving state of the art accuracy, while maintaining real-time detection speed. The general idea behind our work is to create an efficient environment for developing and analyzing novel or optimized approaches in terms of classification, features, usage of prior knowledge and custom strategies for training and detection. In our approach we opt for a standard linear classifier such as Adaboost. Inspired by the integral channel feature approach, we compute variants of LBP and Haar-like features on multiple channels of the input image. Thus, we obtain an elevated number of computationally inexpensive features that capture substantial information. We use an extensive training technique in order to obtain optimal classifier.

We propose a comprehensive framework for object detection with an intuitive modular design and high emphasis on performance and flexibility. Its components are organized by parent-modules, child-modules and auxiliary-modules. The parent-modules contain several child-modules and focus on a general task such as Training or Detection. Child-modules solve more specific tasks, such as feature extraction, training or testing and in most cases require auxiliary-modules. The later have precise intents, for instance computing a color channel transformation or a feature response.

We present two detection configurations. One relies on a single intensively trained detector and the other as a set of specialist detectors.

Our baseline detector uses cascades in order to speed up the classifier. By removing most false positive at first stages, computation time is significantly reduced. Classifier for each cascade is generated using the training approach.

Our contribution is in the form of a hierarchical design of specialized detectors. At first level we use a version of the baseline detector in order to remove irrelevant candidates. At the second level, specialist detectors are defined. These detectors can be either independent or can use third level detectors and cumulate their output. A specialist detector can take the role of solving an exact classification issue, such as sitting pose. In that case it is trained only with data relevant to that task. In some applications, a specialist detector can be trained to perform exceptionally on a specific situation. In this case training samples are adapted to the particularity of the testing, and possibly parts of the testing sets are used for training.

This is a versatile system for object detection that excels in both accuracy and speed. We present a valuable strategy for training and a hierarchy of specialized people detectors for dealing with difficult scenarios. We also propose an interesting feature channel and a method for loosing less detection speed-up. In our approach we build upon the ideas of feature scaling instead of resizing images and of transferring most computations from detection to training, thus achieving real-time performance on VGA resolution.

Figure 25 and Figure 26 illustrate our detections results. Figure 27 shows the performance of our system compared to other. IDIAP detector was used without tuning the parameters.

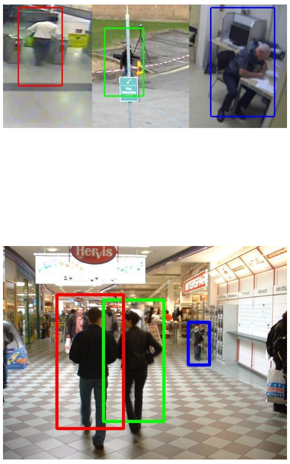


Figure 25. Detection Results

6.15. A Model-based Framework for Activity Recognition of Older People using Multiple sensors

Participants: Carlos -Fernando Crispim Junior, Qiao Ma, Baptiste Fosty, Cintia Corti, Véronique Joumier, Philippe Robert, Alexandra Konig, François Brémond, Monique Thonnat.

keywords: Activity Recognition, Multi-sensor Analysis, Surveillance System, Older people, Frailty assessment

We have been investigating a model-based activity recognition framework for the automatic detection of physical activity tests and instrumental activities of daily living (IADL, *e.g.*, preparing coffee, making a phone call) of older people. The activities are modelled using a constraint-based approach (using spatial, temporal, and *a priori* information of the scene), and a generic ontology based on natural terms which allows medical experts to easily modify the defined activity models. Activity models are organized in a hierarchical structure according to their complexity (Primitive state, Composite State, Primitive Event, and Composite Event). The framework has been tested as a system on the clinical protocol developed by the Memory Center of Nice hospital. This clinical protocol aims at studying how ICTs (Information and Communication Technologies)

Figure 26. PETS Detection Results

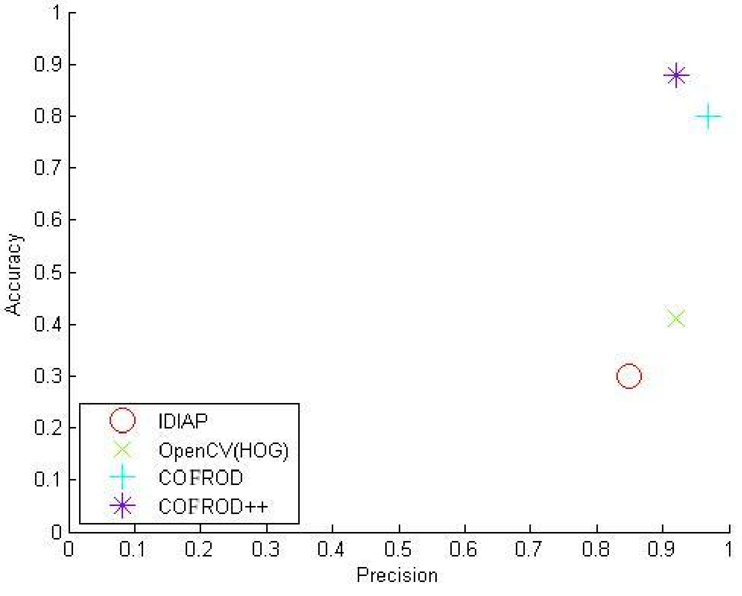


Figure 27. Detection results on the PETS dataset

can provide objective evidence of early symptoms of Alzheimer's disease (AD) and related conditions (like Memory Cognitive Impairment - MCI). The Clinical protocol participants are recorded using a RGB video-camera (8 fps), a RGB-D Camera (Kinect - Microsoft), and an inertial sensor (MotionPod) which allows a multi-sensor evaluation of the activities of the participants in an observation room equipped with home appliances. A study of the use of a multi-sensor monitoring for Patient diagnosis using events annotated by experts has been performed in partnership with CHU-Nice and SMILE team of TAIWAN, and it has shown the feasibility of the use of these sensors for patient performance evaluation and differentiation of clinical protocol groups (Alzheimer's disease and healthy participants) [31] and [40]. The multi-sensor evaluation has used the proposed surveillance system prototype and has been able to detect the full set of physical activities of the scenario 1 of the clinical protocol (e.g., Guided a ctivities : Balance test, Repeated Transfer Test), with a true positive rate of 96.9% to 100% for a set of 38 patients (MCI=19, Alzheimer=9) using data of an ambient camera. An extension of the developed framework has been investigated to handle multiple sensors data in the event modeling. In this new scenario, information from the ambient camera and the inertial sensor worn on the participants chest is used (see Figure 28). The prototype using the extended framework has been tested on the automatic detection of IADLs, and preliminary results points to an average sensitivity of 91% and an average precision of 83.5%. This evaluation has been performed for 9 participants videos (15 min each, healthy: 4, MCI: 5). See [39] for more details. Future work will focus on a learning mechanism to automatic fuse events detected by a set of heterogeneous sensors, and at supporting clinicians at the task of studying differences between the activity profile of healthy participants and early to moderate stage Alzheimer's patients.

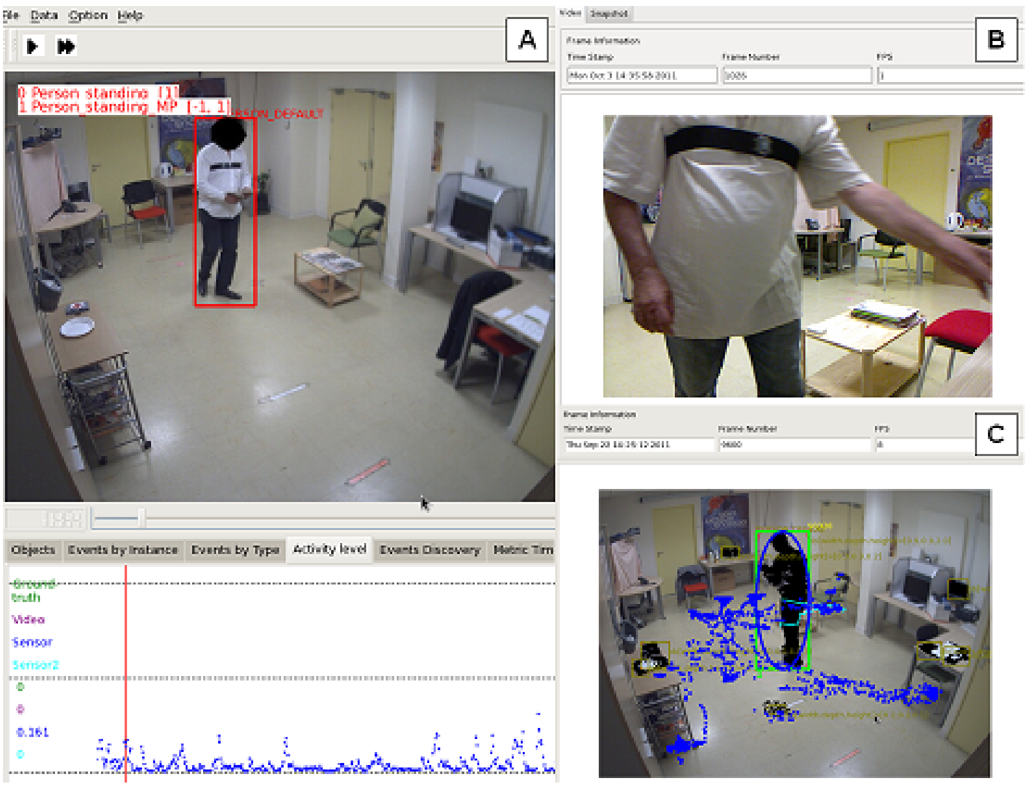


Figure 28. A: Ambient Camera View of Patient Activity. Actimetry captured by the inertial sensor is displayed at the bottom, B: RGB-D Camera View of Patient. The inertial sensor is worn by the patient by an accessory chest strap, C: Trajectory information of Patient Activity during the experimentation.

6.16. Activity Recognition for Older People using Kinect

Participants: Baptiste Fosty, Carlos -Fernando Crispim Junior, Véronique Joumier, Philippe Robert, Alexandra Konig, François Brémond, Monique Thonnat.

keywords: Activity Recognition, RGB-D camera analysis, Surveillance System, Older people, Frailty assessment

Within the context of the Dem@Care project, we have studied the potential of the RGB-D camera (Red Green Blue + Depth) from Microsoft (Kinect) for an activity recognition system developed to extract automatically and objectively evidences of early symptoms of Alzheimer's disease (AD) and related conditions (like Memory Cognitive Impairment - MCI) for older people. This system is designed on a model-based activity recognition framework. Using a constraint-based approach with contextual and spatio-temporal informations of the scene, we have developped activity models related to the physical activity part of the protocol (Scenario 1, guided activities : balance test, walking test, repeated transfers posture between sitting and standing). These models are organized in a hierarchical structure according to their complexity (Primitive state, Composite State, Primitive Event, and Composite Event). This work is an adaptation of the work performed for multi-sensor analysis [39].

Several steps are needed to adapt the processing. We had for example to generate new ground truth, or we had to design new 3D zones of interest according to Kinect point of view and referential (differing from the 2D camera). Moreover, in order to improve the reliability of the results, we had to solve several issues in the processing chain. For instance, Kinect and the detection algorithm provided by OpenNi and Nestk (free libraries) have several limitations which leads to wrong detection of human. We proposed in these cases several solutions like filtering wrong object detections by size (see Figure29 C) or recomputing the height of older people based on their head when wearing black pants (absorption of infrared) (see Figure 29 D).

For the experimentation, we have processed the data recorded for 30 patients. The results are shown in Figure 30 . With a true positive rate of almost 97% and a precision of 94.2%, our system is able to extract most of the activities performed by patients. Then, relevant and objective information can be delivered to clinicians, to assess the patient frailty. For further information on the performance of the detection process, we also generate the results frame by frame, which are shown in Figure 31 . We see there that the performance of the event detection in terms of true positive rate is almost as good as by events (94.5%). Nevertheless, if we focus on the precision, it is lower than previously. This means that we still need to improve detection accuracy of the beginning and the end of an event.

Future work will focus on using the human skeleton to extract finest information on the patient activity and to process more scenarios (semi-guided and free).

6.17. Descriptors of Depth-Camera Videos for Alzheimer Symptom Detection

Participants: Guillaume Charpiat, Sorana Capalnean, Bertrand Simon, Baptiste Fosty, Véronique Joumier.

keywords: Kinect, action description, video analysis

In a collaboration with the CHU hospital of Nice, a dataset of videos was recorded, where elderly are asked by doctors to perform a number of predefined exercises (like walking, standing-sitting, equilibrium test), and recorded with an RGBD camera (Kinect). Our task is to analyze the videos and detect automatically early Alzheimer symptoms, through statistical learning. Here we focus on the 3D depth sensor (no use of the RGB image), and aim at providing action descriptors that are accurate enough to be informative.

During her internship in the Stars team, Sorana Capalnean proposed descriptors relying directly on the 3D points of the scene. First, based on trajectory analysis, she proposed a way to recognize the different physical exercises. Then she proposed, for each exercise, specific descriptors aiming at providing the information asked by doctors, such as step length, frequency and asymmetry for the walking exercise, or sitting speed and acceleration for the second exercise, etc. Problems to deal with included the high level of noise in the 3D cloud of points given by the Kinect, as well as an accurate localization of the floor.

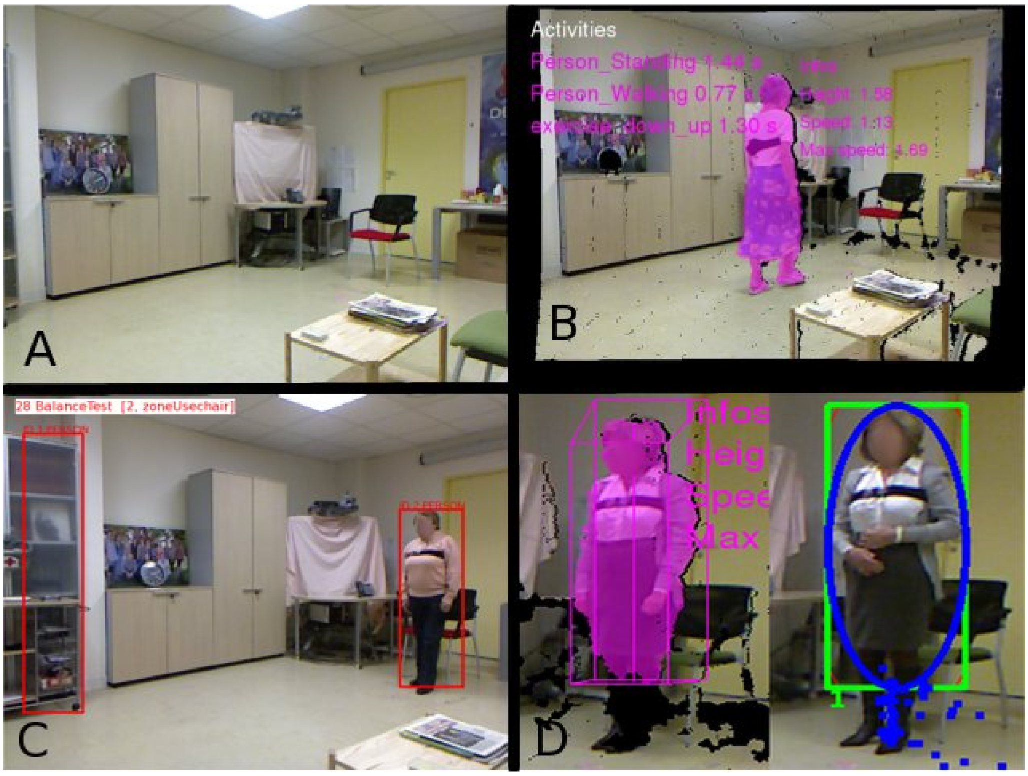


Figure 29. A: RGB-D camera view of the scene, B: 3D representation of the scene with some event detection, C: people detection problem (furniture detected as extra person), D: people detection problem (black clothes not detected).

	Nb GT	Detected		Sensitivity (%)	Precision (%)	Fscore (%)
BalanceTest	30	VP	30	100	100	100
		FP	0			
		FN	0			
WalkingTest_firstAttempt	30	TP	30	100	90,9	95,2
		FP	3			
		FN	0			
WalkingTest_secondAttempt	30	TP	27	90	100	94,7
		FP	0			
		FN	3			
RepeatedTransfersTest	30	TP	30	100	90,9	95,2
		FP	3			
		FN	0			
UpAndGoTest	30	TP	28	93,3	90,3	91,8
		FP	3			
		FN	2			
TOTAL	**150**	**TP**	**145**	**96,6**	**94,2**	**95,4**
		FP	**9**			
		FN	**5**			

Figure 30. Results by events (GT = ground truth, TP = true positive, FP = false positive, FN = false negative)

	Nb GT	Detected		Sensitivity (%)	Precision (%)	Fscore (%)
BalanceTest	37235	TP	37216	99,9	70,8	82,9
		FP	15322			
		FN	19			
WalkingTest_firstAttempt	1537	TP	849	55,2	94,1	69,6
		FP	53			
		FN	688			
WalkingTest_secondAttempt	1425	TP	857	60,1	62,6	61,3
		FP	513			
		FN	568			
RepeatedTransfersTest	5190	TP	4496	86,6	94,6	90,4
		FP	257			
		FN	694			
UpAndGoTest	3345	TP	2670	79,8	86,2	82,9
		FP	428			
		FN	675			
TOTAL	**48732**	**TP**	**46088**	**94,5**	**73,6**	**82,8**
		FP	**16573**			
		FN	**2644**			

Figure 31. Results by events (GT = ground truth, TP = true positive, FP = false positive, FN = false negative)

During his internship, Bertrand Simon proposed other kinds of descriptors, based on the articulations of the human skeleton given by OpenNI. These articulations are however very noisy too, so that a pre-filtering step of the data in time had to be performed. Various coordinate systems were studied, to reach the highest robustness. The work focused not only on descriptors but also on metrics suitable to compare gestures (in the phase space as well as in the space of trajectories). See figure 32 for an example.

These descriptors are designed to be robust to camera noise and to extract the relevant information from the videos; however their statistical analysis still remains to be done, to recognize Alzheimer symptoms during the different exercises.

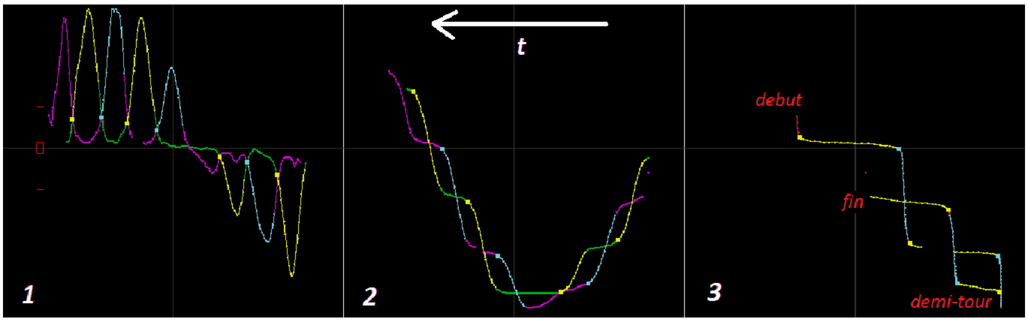


Figure 32. Curves obtained during a person's walk in backward then forward directions. Purple and blue curves stand for the right foot, while green and yellow ones stand for the left one. Graph 1 shows the speeds as a function of time; Graph 2 shows the locations as a function of time; Graph 3 shows the location of the right foot as a function of the location of the left foot.

6.18. Online Activity Learning from Subway Surveillance Videos

Participants: Jose-Luis Patino Vilchis, Abhineshwar Tomar, François Brémond, Monique Thonnat.

Keywords: Activity learning, clustering, trajectory analysis, subway surveillance

This work provides a new method for activity learning from subway surveillance videos. This is achieved by learning the main activity zones in the observed scene by taking as input the trajectories of detected mobile objects. This provides us the information on the occupancy of the different areas of the scene. In a second step, these learned zones are employed to extract people activities by relating mobile trajectories to the learned zones, in this way, the activity of a person can be summarised as the series of zones that the person has visited. If the person resides in the single zone this activity is also classified as a standing. For the analysis of the trajectory, a multiresolution analysis is set such that a trajectory is segmented into a series of tracklets based on changing speed points thus extracting the information when people stop to interact with elements of the scene or other people. Starting and ending tracklet points are fed to an advantageous incremental clustering algorithm to create an initial partition of the scene. Similarity relations between resulting clusters are modelled employing fuzzy relations. A clustering algorithm based on the transitive closure calculation of the fuzzy relations easily builds the final structure of the scene. To allow for incremental learning and update of activity zones (and thus people activities), fuzzy relations are defined with online learning terms. The approach is tested on the extraction of activities from the video recorded at one entrance hall in the Torino (Italy) underground system. Figure 33 presents the learned zones corresponding to the analyzed video. To test the validity of the activity extraction a one hour video was annotated with activities (corresponding to each trajectory) according to user defined ground-truth zones. After the comparison, following results were obtained: TP:26, FP:3, FN:1, Precision:0.89, Sensitivity:0.96. This work is published in [43].

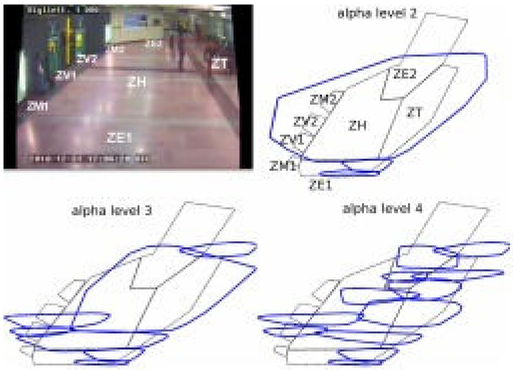


Figure 33. Left top panel: Original underground scene observed by the camera with user-defined areas delimiting the scene. Remaining panels: Learned zones in a 3D top view. They correspond to activity areas as discovered with our algorithm. Different granularity levels allow understanding the activity with different resolutions

6.19. Automatic Activity Detection Modeling and Recognition: ADMR

Participants: Guido-Tomas Pusiol, François Brémond.

This year a new Ph.D. thesis has been defended [30]. The main objective of the thesis is to propose a complete framework for the automatic activity discovery, modeling and recognition using video information. The framework uses perceptual information (e.g. trajectories) as input and goes up to activities (semantics). The framework is divided into five main parts:

1. We break the video into chunks to characterize activities. We propose different techniques to extract perceptual features from the chunks. This way, we build packages of perceptual features capable of describing activity occurring in small periods of time.
2. We propose to learn the video contextual information. We build scene models by learning salient perceptual features. The models end up containing interesting scene regions capable of describing basic semantics (i.e. region where interactions occur).
3. We propose to reduce the gap between low-level vision information and semantic interpretation, by building an intermediate layer composed of Primitive Events. The proposed representation for primitive events aims at describing the meaningful motions over the scene. This is achieved by abstracting perceptual features using contextual information in an unsupervised manner.
4. We propose a pattern-based method to discover activities at multiple resolutions (i.e. activities and sub-activities). Also, we propose a generative method to model multi-resolution activities. The models are built as a flexible probabilistic framework easy to update.
5. We propose an activity recognition method that finds in a deterministic manner the occurrences of modelled activities in unseen datasets. Semantics are provided by the method under user interaction. All this research work has been evaluated using real datasets of people living in an apartment (home-care application) and elder patients in a hospital.

The work has also been evaluated for other types of applications such as sleeping monitoring. For example, Figure 34 display the results of the activity discovery method during 6 hours (left to right) applied to the center of mass (3D) of a tracked sleeping person. The colored segments represent hierarchical (bottom-up is finer-coarse) discovered activity which matches with sleeping postural movements. The segments have similar color when postural movements are similar. For example, the segment (j) is the only time the person sleeps upside down. Also, health professionals analysed the results claiming that the segments corresponds to normal sleeping cycle, where low motion is noticed at the beginning of the sleep and more motion is shown when the person have a lighter sleep when starts waking up.

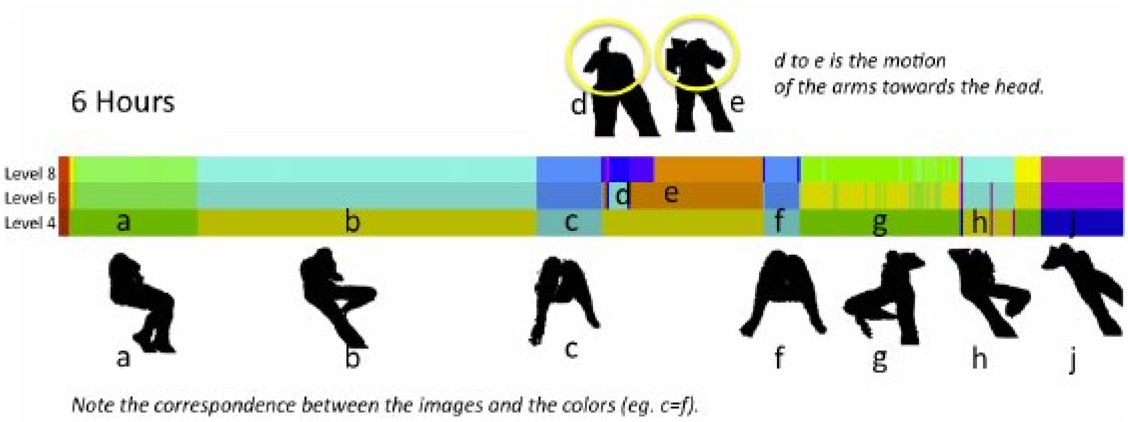


Figure 34. Results of the activity discovery method during 6 hours

6.20. SUP Software Platform

Participants: Julien Gueytat, Baptiste Fosty, Anh tuan Nghiem, Leonardo Rocha, François Brémond.

Our team focuses on developing Scene Understanding Platform (SUP) (see section 5.1). This platform has been designed for analyzing a video content. SUP is able to recognize simple events such as 'falling', 'walking' of a person. We can easily build new analyzing system thanks to a set of algorithms also called plugins. The order of those plugins and their parameters can be changed at run time and the result visualized. This platform has many more advantages such as easy serialization to save and replay a scene, portability to Mac, Windows or Linux, ... All those advantages are available since we are working together with the software developers team DREAM. Many Inria teams are pushing together to improve a common Inria development toolkit DTK. Our SUP framework is one of the DTK-like framework developed at Inria.
Currently, we have fully integrated OpenCV library with SUP and the next step is to integrate OpenNI to get depth map processing algorithms from PrimeSense running in SUP. Updates and presentations of our framework can be found on our team website http://team.inria.fr/stars. Detailed tips for users are given on our Wiki website http://wiki.inria.fr/stars and sources are hosted thanks to the new Source Control Management tool.

6.21. Qualitative Evaluation of Detection and Tracking Performance

Participants: Swaminathan Sankaranarayanan, François Brémond.

We study an evaluation approach for detection and tracking systems. Given an algorithm that detects people and simultaneously tracks them, we evaluate its output by considering the complexity of the input scene. Some videos used for the evaluation are recorded using the Kinect sensor which provides for an automated ground truth acquisition system. To analyse the algorithm performance, a number of reasons due to which an algorithm might fail is investigated and quantified over the entire video sequence. A set of features called Scene Complexity measures are obtained for each input frame. The variability in the algorithm performance is modelled by these complexity measures using a polynomial regression model. From the regression statistics, we show that we can compare the performance of two different algorithms and also quantify the relative influence of the scene complexity measures on a given algorithm. This work has been published in [44].

6.22. Model-Driven Engineering and Video-surveillance

Participants: Sabine Moisan, Jean-Paul Rigault, Luis-Emiliano Sanchez.

keywords: Feature Model Optimization, Software Metrics, Requirement specification, Component-based system, Dynamic Adaptive Systems, Model-Driven Engineering, Heuristic Search, Constraint Satisfaction Problems

The domain of video surveillance (VS) offers an ideal training ground for Software Engineering studies, because of the huge variability in both the surveillance tasks and the video analysis algorithms [41]. The various VS tasks (counting, intrusion detection, tracking, scenario recognition) have different requirements. Observation conditions, objects of interest, device configuration... may vary from one application to another. On the implementation side, selecting the components themselves, assembling them, and tuning their parameters to comply with context may lead to a multitude of variants. Moreover, the context is not fixed, it evolves dynamically and requires run time adaptation of the component assembly.

Our work relies on Feature Models, a well-known formalism to represent variability in software systems. This year we have focused on an architecture for run time adaptation and on metrics to drive dynamic architecture changes.

6.22.1. Run Time Adaptation Architecture

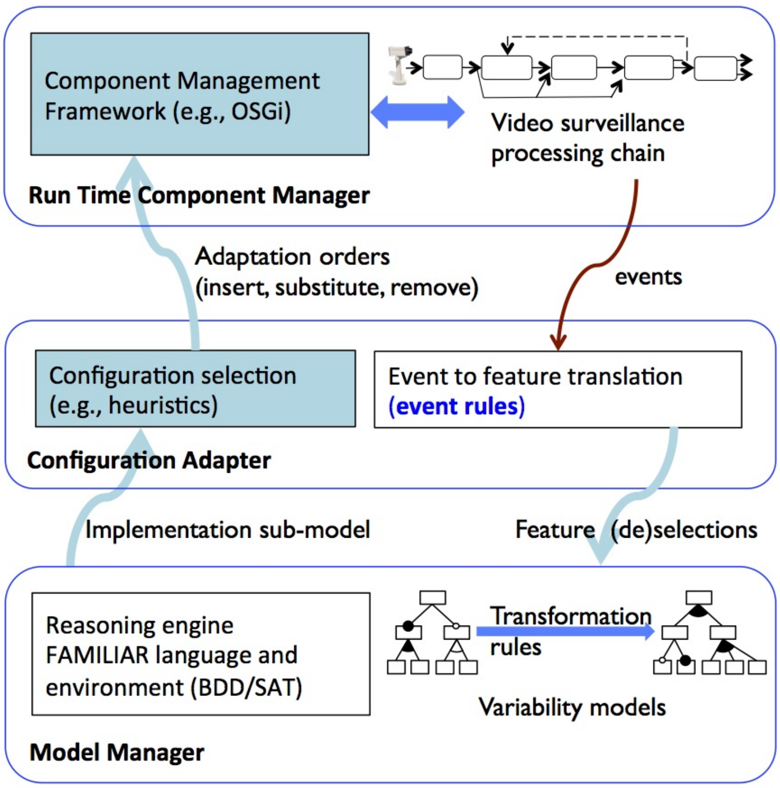


Figure 35. Architecture of an adaptive system using models at run time. (In light blue the elements studied this year.)

The architecture of the run time system (also used for initialization at deployment time) is based on three collaborating modules as shown in Figure 35 . A *Run Time Component Manager* (RTCM) cooperates with the low levels (to manage the software components and capture events) and applies configuration changes. A *Configuration Adapter* (CA) receives events from the RTCM, and propagates them as features into the models to obtain a new configuration. The *Model Manager* (MM) embeds a specialized scripting language for Feature Models (FAMILIAR [52], [53][1]) to manage the representation of the two specialized feature models and applies constraints and model transformations on them. The *Model Manager* produces new component configurations (a model specialization) that it sends to the CA. At its turn, the CA selects one single configuration (possibly using heuristics) and converts it into component operations to be applied by the RTCM.

This year we first finalized the interface between the Model Manager and the Configuration Adapter. On one hand, we transform the feature models obtained from FAMILIAR into C++ representations enriched with software component information. On the other hand, we dynamically transform context change events into requests to FAMILIAR.

Second, we searched for a suitable technology for handling components in the Run Time Component Manager. OSGi is an adequate de facto standard but it is mainly available in the Java world. However we could find a C++ implementation, complete enough for our needs (SOF, Service Oriented Framework [65]). However, SOF has to be completed to adjust to the needs of our end users who are the video system developers. Thus, we are currently building a multi-threaded service layer on top of SOF, easy to use and hiding most of the nitty-gritty technical details of thread programming and SOF component manipulation. This layer provides end users with a set of simple patterns and allow them to concentrate only on the code of video services (such as acquisition, segmentation, tracking...).

As a matter of feasability study we are building an experimental video self adaptive system based on the afore mentionned architecture. Software components are implemented with the OpenCV library. In the final system, feature models and software components continuously interact in real time, modifying the whole system in response to changes in its environment.

6.22.2. Metrics on Feature Models to Optimize Configuration Adaptation at Run Time

As shown on figure 35 , the Configuration Adapter has to set up a suitable component configuration of the run time system. For this, each time the context changes, it receives a set of valid configurations (a feature *sub*-model) from the Model Manager. In most cases, this set contains more than one configuration. Of course, only one configuration can be applied at a given time and the problem is to select the “best” one. Here, “best” is a trade-off between several non-functional aspects: performance, quality of service, time cost for replacing the current configuration, etc.

It is thus necessary to rank the configurations. Our approach is to define metrics suitable for comparing configurations. Then the problem comes down to the widely studied problem of *Feature model optimization* [55]. This problem is known to be an intractable combinatorial optimization problem in general.

We started with a study of the state of the art: metrics for general graphs as well specific to feature models, optimization and requirement specification on feature models... We obtained a structured catalog of quality and feature model metrics. Then we selected solutions based on heuristic search algorithms using quality and feature model metrics. We thus propose several strategies and heuristics offering different properties regarding optimality of results and execution efficiency.

These strategies and heuristics have been implemented, tested, and analyzed using random generated feature models. We got empirical measures about their properties, such as completeness, optimality, time and memory efficiency, scalability... This allows us to compare the performance of the different algorithms and heuristics, and to combine them in order to achieve a good trade-off between optimality and efficiency. Finally, the proposed algorithms have been introduced as part of the Configuration Adapter module.

[1] FAMILIAR has been developed at the I3S laboratory by the Modalis team.

This work is quite original from several aspects. First, we did not find any study using heuristic search algorithms for solving the feature optimization problem. Most studies apply Artificial Intelligence techniques such as CSP solvers, planning agents, genetic algorithms... Second, we do not restrict to the optimization of linear objective functions, but we also address non-linear ones allowing us to take into account a broader set of criteria. Among the possible criteria we consider quality of service of components, their performance, their set up delay, the cost of their replacement, etc. Finally, we apply our metrics at run time whereas most studies consider metrics only for static analysis of feature models.

Currently, we are still working on new variants of the search algorithms and new heuristics relying on techniques proposed in the domains of heuristic search and constraint satisfaction problems.

6.23. Synchronous Modelling and Activity Recognition

Participants: Annie Ressouche, Sabine Moisan, Jean-Paul Rigault, Daniel Gaffé.

6.23.1. Scenario Analysis Module (SAM)

To generate activity recognition systems we supply a scenario analysis module (SAM) to express and recognize complex events from primitive events generated by SUP or other sensors. In this framework, this year we focus on recognition algorithm improvement in order to face the problem of large number of scenario instances recognition.

The purpose of this research axis is to offer a generic tool to express and recognize activities. Genericity means that the tool should accommodate any kind of activities and be easily specialized for a particular framework. In practice, we propose a concrete language to specify activities in the form of a set of scenarios with temporal constraints between scenarios. This language allows domain experts to describe their own scenario models. To recognize instances of these models, we consider the activity descriptions as synchronous reactive systems [76] and we adapt usual techniques of synchronous modelling approach to express scenario behaviours. This approach facilitates scenario validation and allows us to generate a recognizer for each scenario model.

In addition, we have completed SAM in order to address the life cycle of scenario instances. For a given scenario model there may exist several (possibly many) instances at different evolution states. These instances are created and deleted dynamically, according to the input event flow. The challenge is to manage the creation/destruction of this large set of scenario instances efficiently (in time and space), to dispatch events to expecting instances, and to make them evolve independently. To face this challenge, we introduced in the generation of the recognition engine, the expected events of the next step. This avoids to run the engine automatically with events that are not relevant for the recognition process. Indeed, we relied on Lustre [66] synchronous language to express the automata semantics of scenario models as Boolean equation systems. This approach was successful and shows that we can consider a synchronous framework to generate validated scenario recognition engines. This year, in order to improve efficiency (and to tackle the real time recognition problem), we begin to rely on CLEM (see section 6.23.2) toolkit to generate such recognition engines. The reason is threefold: (1) CLEM is becoming a mature synchronous programming environment; (2) we can use the CLEM compiler to build our own compiler; (3) CLEM supplies the possibility of using NuSMV [61] model checker, which is more powerful than the Lustre model-checker. Moreover, thanks to CLEM compiler into Boolean equation systems, we can compute the expected events of the next instant on the fly, by propagation of information related to the current instant.

6.23.2. The clem Workflow

This research axis concerns the theoretical study of a synchronous language LE with modular compilation and the development of a toolkit (see Figure 9) around the language to design, simulate, verify and generate code for programs. The novelty of the approach is the ability to manage both modularity and causality. This year, we mainly work on theoretical aspects of CLEM.

First, synchronous language semantics usually characterizes each output and local signal status (as present or absent) according to input signal status. To reach our goal, we defined a semantics that translates LE programs into equation systems. This semantics bears and grows richer the knowledge about signals and is never in contradiction with previous deduction (this property is called constructiveness). In such an approach, causality turns out to be a scheduling evaluation problem. We need to determine all the partial orders of equation systems and to compute them, we consider a 4-valued algebra to characterize the knowledge of signal status (unknown, present, absent, overknown). Previously, we relied on 4-valued Boolean algebra [19], [20] which defines the negation of unknown as overknown. The advantage of this way is to benefit from Boolean algebras laws to compute equation system solutions. The drawback concerns signal status evaluation which does not correspond to usual interpretation (not unknown = unknown and not overknown = overknown). To avoid this drawback, we study other kinds of algebras well suited to define synchronous languages semantics. In [49], we choose an algebra which is a bilattice and we show that it is well suited to solve our problem. It is a new application of general bilattice theory [64]. But, the algebra we defined is no more a Boolean algebra, but we prove (always in [49]), that the main laws of Boolean algebras hold as distributivity laws, associativity laws, idempotence laws, etc. After compilation, signals have to be projected into Boolean values. Bilattice theory offers an isomorphism between 4-valued status and pair of Boolean.

Second, the algorithm which computes partial orders relies on the computation of two dependency graphs: the upstream (downstream) dependency graph computes the dependencies of each variable of the system starting from the input (output) variables. Inputs (resp. outputs) have date 0 and the algorithm recursively increases the dates of nodes in the upstream (resp downstream) dependencies graph. Hence, the algorithm determines an *earliest* date and a *latest* date for equation system variables. Moreover, we can compute the dates of variables of a global equation system starting from dates already computed for variables which were inputs and outputs in a sub equation system corresponding to a sub program of the global program[2]. This way of compiling is the corner stone of our approach [20]. We defined two approaches to compute all the valid partial orders of equation systems, either applying critical path scheduling technique (CPM) [3] or applying fix point theory: the vector of earliest (resp. latest) dates can be computed as the least fix point of a monotonic increasing function. This year we have proved that we can compute dates either starting from a global equation system or considering equation system where some variables are abstracted (i.e they have no definition) and whose dates have been already computed. To achieve the demonstration, we rely on an algebraic characterization of dates and thanks to uniqueness property of least fix points, we can deduce that the result is the same for a global equation systems as for its abstraction. We are in the process of publishing this result. From an implementation point of view, we use CPM approach to implement our scheduling algorithm since it is more efficient than fix point consideration. Of course both ways yield the same result. Indeed, fix point approach is useful for a theoretical concern.

6.23.3. Multiple Services for Device Adaptive Platform for Scenario Recognition

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 5.1) with a service oriented middleware approach dealing with dynamic evolutions of system infrastructure. The Rainbow team (Nice-Sophia Antipolis University) proposes a component-based adaptive middleware (WComp [85], [84], [68]) to dynamically adapt and recompose assemblies of components. These operations must obey the "usage contract" of components. The existing approaches don't really ensure that this usage contract is not violated during application design. Only a formal analysis of the component behaviour models associated with a well sound modelling of composition operation may guarantee the respect of the usage contract.

The approach we adopted introduces in a main assembly, a synchronous component for each sub assembly connected with a critical component. This additional component implements a behavioural model of the critical component and model checking techniques apply to verify safety properties concerning this critical component. Thus, we consider that the critical component is validated.

[2] these variables are local in the global equation system
[3] http://pmbook.ce.cmu.edu/10_Fundamental_Scheduling_Procedures.html

To define such synchronous component, user can specify a synchronous component per sub assembly corresponding to a concern and compose the synchronous components connected with the same critical component in order to get an only synchronous component. Thus, we supply a *composition under constraints* of synchronous components and we proved that this operation preserves already separately verified properties of synchronous components [79], [78].

The main challenge of this approach is to deal with the possibly very large number of constraints a user must specify. Indeed, each synchronous monitor has to tell how it combines with other, then we get a combinatorial number of constraints with respect to the number of synchronous monitors and inputs of the critical component. To tackle this problem, we replace the effective description of constraints by a generic specification of them in the critical component. But, we must offer a way to express these generic constraints. Then, each synchronous component has a synchronous controller, which is the projection of the generic constraints on its output set. The global synchronous component is the synchronous parallel composition of all basic components and their synchronous controllers. Moreover, according to synchronous parallel composition features, the property preservation result we have still hold.

TEXMEX Project-Team

6. New Results

6.1. Description of multimedia content

6.1.1. Face Recognition

Participants: Thanh Toan Do, Ewa Kijak.

Face recognition is an important tool for many applications like video analysis. We addressed the problem of faces representation and proposed a weighted co-occurrence Histogram of Oriented Gradient as facial representation. The approach was evaluated on two typical face recognition datasets and has shown an improvement of the recognition rate over state of the art methods [31].

6.1.2. Violent scene detection

Participants: Guillaume Gravier, Patrick Gros, Cédric Penet.

Joint work with Technicolor.

We have worked on multimodal detection of violent scenes in Hollywood movies, in collaboration with Technicolor. Two main directions were explored. On the one hand, we investigated different kinds of Bayesian network structure learning algorithms for the fusion of multimodal features [49]. On the other hand, we studied the use of audio words for the detection of violent related events—gunshots, screams and explosions—in the soundtrack, demonstrating the benefit of product quantization and multiple words representations for increased robustness to variability between movies.

6.1.3. Text detection in videos

Participants: Khaoula Elagouni, Pascale Sébillot.

Joint work with Orange Labs.

Texts embedded in videos often provide high level semantic clues that can be used in several applications and services. We thus aim at designing efficient Optical Character Recognition (OCR) systems able to recognize these texts. In 2012, we proposed a novel approach that avoids the difficult step of character segmentation. Using a multi-scale scanning scheme, texts extracted from videos are first represented by sequences of features learnt by a convolutional neural network. The obtained representations fed a connectionist recurrent model, that relies on the combination of a BLSTM and a CTC connectionist classification model, specifically designed to take into account dependencies between successive learnt features and to recognize texts. The proposed video OCR, evaluated on a database of TV news videos, achieves very high recognition rates (character recognition rate: 97%; word recognition rate: 87%). Experiments also demonstrate that, for our recognition task, learnt feature representations perform better than standard hand-crafted features ([34]). We also carried out a comparison between two of our previous text recognition methods, one relying on a character segmentation step, the other one avoiding it by using a graph model, both on natural scene texts and embedded texts, highlighting the advantages and the limits of each of them. This work is submitted to the journal IJDAR.

6.1.4. Automatic speech recognition

Participants: Guillaume Gravier, Bogdan Ludusan.

This work was partly performed in the context of the Quaero project and the ANR project Attelage de Systèmes Hétérogènes (ANR-09-BLAN-0161-03), in collaboration with the METISS project-team.

In a multimedia context, automatic speech recognition (ASR) provides semantic access to multimedia but faces robustness issues due to the diversity of media sources. To increase robustness, we explore new paradigms for speech recognition based on collaborative decoding and phonetically driven decoding. We investigated mechanisms for the interaction of multiple ASR systems, exchanging linguistic information in a collaborative setting [15]. Following the same idea, we proposed phonetically driven decoding algorithms where the ASR system makes use of phonetic landmarks (place and manner of articulation, stress) to bias and prune the search space [65], [70]. In particular, we proposed a new classification approach to broad phonetic landmark detection [69].

6.2. Large scale indexing and classification techniques

6.2.1. Image retrieval and classification

Participants: Rachid BenMokhtar, Jonathan Delhumeau, Patrick Gros, Mihir Jain, Hervé Jégou, Josip Krapac.

This work was partially done in collaboration with Matthijs Douze and Cordelia Schmid (LEAR), Florent Perronnin and Jorge Sanchez (Xerox), Patrick Pérez (Technicolor) and Ondrej Chum (CVUT Prague). It was partly done in the context of the Quaero project.

Our work on very large scale image search has addressed [14] the joint optimization of three antinomic criterions: speed, memory resources and search quality. We have considered techniques aggregating local image descriptors into a vector and show that the Fisher kernel achieves better performance than the reference bag-of-visual words approach for any given vector dimension. The joint optimization of dimensionality reduction with indexing allowed us to obtain a precise vector comparison as well as a compact representation. The evaluation shows that the image representation can be reduced to a few dozen bytes while preserving high accuracy. Searching a 100 million image dataset takes about 250 ms on one processor core.

This work has been further improved [45] by modifying the way the similarity between images is computed, in particular we have shown that whitening is an effective way to fully exploit multiple vocabularies along with bag-of-visual words and VLAD representations.

We have also considered the problem of image classification, which goal is to produce a semantic representation of the images in the form of text labels reflecting the object categories contained in the images. We have proposed a technique derived from a matching system [44] based on Hamming Embedding and a similarity space mapping. The results outperform the state-of-the-art among matching systems such as NBNN. On some datasets such as Caltech-256, our results compare favorably to the best techniques, namely the Fisher vector representation.

6.2.2. Intensive use of SVM for text mining and image mining

Participants: Thanh Nghi Doan, François Poulet.

Following our previous work on large scale image classification [58], we have developed a fast and efficient framework for large scale image classification. Most of the state of the art approaches use a linear SVM (eg LIBLINEAR) for the training task. Another solution can be to use the new Power Mean SVM (PmSVM) with power mean kernel functions that can solve a binary classification problem with millions of examples and tens of thousands of dense features in a few seconds (excluding the time to read the input files). We are working on a parallel version of this algorithm and trying to deal with unbalanced datasets: in ImageNet1000 dataset, there are 1,000 classes, this is a very unbalanced classification task so we use a balanced bagging parallel algorithm. The time needed to perform the training task on ImageNet1000 was almost 1 day with the original PmSVM algorithm and 2.5 days for LIBLINEAR, we achieve it within 10 min and with a relative precision increase of more than 20%. We are currently working to reduce the RAM needed to perform the task (today 30GB).

6.2.3. Audio indexing

Participants: Jonathan Delhumeau, Guillaume Gravier, Patrick Gros, Hervé Jégou.

This work was done in the context of the Quaero project.

Our new Babaz audio search system [46] aims at finding modified audio segments in large databases of music or video tracks. It is based on an efficient audio feature matching system which exploits the reciprocal nearest neighbors to produce a per-match similarity score. Temporal consistency is taken into account based on the audio matches, and boundary estimation allows the precise localization of the matching segments. The method is mainly intended for video retrieval based on their audio track, as typically evaluated in the copy detection task of Trecvid evaluation campaigns. The evaluation conducted on music retrieval shows that our system is comparable to a reference audio fingerprinting system for music retrieval, and significantly outperforms it on audio-based video retrieval, as shown by our experiments conducted on the dataset used in the copy detection task of the Trecvid'2010 campaign, which was used as an external evaluation in the Quaero project.

6.2.4. Approximate nearest neighbor search with compact codes

Participants: Teddy Furon, Hervé Jégou.

This work was done in collaboration with the Metiss project team (Anthony Bourrier and Rémi Gribonval). It was partly done in the context of the Quaero project.

Following recent works on Hamming Embedding techniques, we proposed [47] a binarization method that aim at addressing the problem of nearest neighbor search for the Euclidean metric by mapping the original vectors into binary vectors ones, which are compact in memory, and for which the distance computation is more efficient. Our method is based on the recent concept of anti-sparse coding, which exhibits here excellent performance for approximate nearest neighbor search. Unlike other binarization schemes, this framework allows, up to a scaling factor, the explicit reconstruction from the binary representation of the original vector. We also show that random projections which are used in Locality Sensitive Hashing algorithms, are significantly outperformed by regular frames for both synthetic and real data if the number of bits exceeds the vector dimensionality, i.e., when high precision is required.

Another aspect we have investigated in this line of research is the problem of efficient nearest neighbor search for arbitrary kernels. For this purpose, we have combined [76] the product quantization technique [4] with explicit embeddings, and showed that this solution significantly outperforms the state-of-the-art technique designed for arbitrary kernels, such as Kernelized Locality Sensitive Hashing. In addition, we have proposed a variant to perform the exact search.

6.2.5. Indexing and searching large image collections with map-reduce

Participants: Laurent Amsaleg, Gylfi Gudmundsson.

This work was done in the context of the Quaero project.

Most researchers working on high-dimensional indexing agree on the following three trends: (i) the size of the multimedia collections to index are now reaching millions if not billions of items, (ii) the computers we use every day now come with multiple cores and (iii) hardware becomes more available, thanks to easier access to Grids and/or Clouds. This work shows how the Map-Reduce paradigm can be applied to indexing algorithms and demonstrates that great scalability can be achieved using Hadoop, a popular Map-Reduce-based framework. Dramatic performance improvements are not however guaranteed a priori: Such frameworks are rigid, they severely constrain the possible access patterns to data and the RAM memory has to be shared. Furthermore, algorithms require major redesign, and may have to settle for sub-optimal behavior. The benefits, however, are numerous: Simplicity for programmers, automatic distribution, fault tolerance, failure detection and automatic re-runs and, last but not least, scalability. We report our experience of adapting a clustering-based high-dimensional indexing algorithm to the Map-Reduce model, and of testing it at large scale with Hadoop as we index 30 billion SIFT descriptors. We draw several lessons from this work that could minimize time, effort and energy invested by other researchers and practitioners working in similar directions.

6.2.6. Vectorization

Participant: Vincent Claveau.

The vectorization principle allows the description of any object in a vector space based on its similarity with pivots objects. During the last years, we have shown that such a technique can be successfully used for Information Retrieval or Topic Segmentation. This year, TexMex has demonstrated how it can be used in a pure data-mining framework by participating to the JRS2012 framework. The task proposed was a high-dimensional and multi-class machine learning problem. Our approach, based on a simple kNN using vectorization has proved its interest, since it was ranked in top-methods while requiring no training phase nor complex setting.

6.3. Security of media

6.3.1. Security of content based image retrieval

Participants: Laurent Amsaleg, Thanh Toan Do, Teddy Furon, Ewa Kijak.

The performance of Content-Based Image Retrieval Systems (CBIRS) is typically evaluated via benchmarking their capacity to match images despite various generic distortions such as cropping, rescaling or Picture in Picture (PiP) attacks, which are the most challenging. Distortions are made in a very generic manner, by applying a set of transformations that are completely independent from the systems later performing recognition tasks. Recently, studies have shown that exploiting the finest details of the various techniques used in a CBIRS offers the opportunity to create distortions that dramatically reduce the recognition performance [30]. Such a *security perspective* is taken in our work. Instead of creating generic PiP distortions, we have proposed a creation scheme able to delude the recognition capabilities of a CBIRS that is representative of state of the art techniques as it relies on SIFT, high-dimensional k-nearest neighbors searches and geometrical robustification steps. We have ran experiments using 100,000 real-world images confirming the effectiveness of these security-oriented PiP visual modifications [29]. This work goes together with the completed PhD of Thanh-Toan Do [8].

6.3.2. The concept of effective key length in watermarking

Participant: Teddy Furon.

Whereas the embedding distortion, the payload and the robustness of digital watermarking schemes are well understood, the notion of security is still not completely well defined. The approach proposed in the last five years is too theoretical and solely considers the embedding process, which is half of the watermarking scheme. In collaboration with Patrick BAS (CNRS, Ecole Centrale de Lille), we propose a new measure of watermarking security. This concept is called the *effective key length*, and it captures the difficulty for the adversary to get access to the watermarking channel: The adversary proposes a test key and the security is measured as the probability that this test key grants him the watermarking channel (he succeeds to decode hidden messages).

This new methodology is applied to the most wide spread watermarking schemes where theoretical and practical computations of the effective key length are proposed: Zero-bit ‘Broken Arrows’ technique [22], spread spectrum (SS) based schemes (like additive SS, improved SS, and correlation aware SS) [23], and quantization index modulation (QIM) scheme (like Distortion Compensated QIM) [38]. A journal article about this new concept has been submitted to IEEE Trans. on Information Forensics and Security. The keystone of the approach is the evaluation of a security level to the estimation of a probability. Experimental protocols using rare event probability estimator allow good evaluation of this quantity. The soundness of this latter estimator has been theoretically proven in [11] (collaboration with Inria team-project ALEA and ASPI).

6.3.3. A practical joint decoder for active fingerprinting

Participant: Teddy Furon.

This work deals with active fingerprinting, a.k.a. traitor tracing. A robust watermarking technique embeds the user’s codeword into the content to be distributed. When a pirated copy of the content is scouted, the watermark decoder extracts the message, which identifies the dishonest user. However, there might exist a group of dishonest users, so called collusion, who mix their personal versions of the content to forge the pirated copy. The extracted message no longer corresponds to the codeword of one user, but is a mix of several codewords. The decoder aims at finding back some of these codewords to identify the colluders, while avoiding accusing innocent users.

This work follows our breakthrough on Tardos code joint decoding, mentioned in last year's activity report, and whose journal version has been published this year in [16]. Information theory proves that a joint decoder computing scores for pairs, triplets, or in general t-tuples of users is more powerful than single decoders working with scores for single users. However, nobody did try them for large scale setups since the number of t-tuples is in $O(n^t)$. In practical scenarios, n is at least 10,000 and t is around 10, which implies the computation of $\sim 10^{40}$ scores. Last year, we were the first team to design an approximate joint decoder. If its complexity was well under control (in $O(n)$), its iterative structure was much intricate.

Our new design of joint decoder is based on the Monte-Carlo Markov Chain method. It is a simpler iterative process allowing us to sample collusion subsets according to the A Posteriori distribution. Then, the probability that user j is guilty is empirically evaluated over this sample, and threshold to yield a reliable decision. This work has been done under a collaboration with Inria team-project ASPI, and published in [39].

6.4. Multimedia content structuring

6.4.1. Motif discovery

Participants: Guillaume Gravier, Hervé Jégou, Anh Phuong Ta, Wanlei Zhao.

This work was done in the context of the Quaero project.

We have pursued our work on unsupervised discovery of repeating motifs in multimedia data along three directions:

- Discovery of multiple recurrent audio-visually consistent sequences: We proposed two unsupervised approaches to automatically detect multiple structural events in videos using audio and visual modalities. Both approaches rely on cross-modal cluster analysis techniques to directly define events from the data without any prior assumption [51], [52].
- Large-scale unsupervised discovery of near-duplicate shots in TV streams: We developed an efficient method with little a priori knowledge which relies on a product k-means quantizer to efficiently produce hash keys adapted to the data distribution of the frame descriptors. This hashing technique combined with a temporal consistency check allows the detection of meaningful repetitions in TV streams [54].
- Audio motif discovery: This joint work with the METISS project-team extends the generic audio motif discovery method developed in the Ph. D. thesis of Armando Muscariello [17]. We developed an efficient implementation, which will be made publicly available. The software was benchmarked on near duplicate audio motif discovery in the framework of the Quaero project.

6.4.2. Stream labeling for TV structuring

Participants: Vincent Claveau, Guillaume Gravier, Patrick Gros, Emmanuelle Martienne, Abir Ncibi.

In this application, we focus on the problem of labeling the segments of a TV stream according to their types (*e.g.*,, programs, commercial breaks, sponsoring, ...). During this year, we performed an in-depth analysis of the use of Conditional Random Fields (CRF) for our task. In particular, we studied:

- how sequentiality is modeled with the CRF;
- the links with other probabilistic graphical techniques (HMM, MEMM...);
- the robustness of the approach when dealing with few training data or few features;

The use of this model for semi-supervised and unsupervised learning are under study. We also studied the use of very simple descriptors (simple shot lengths, and use of global image descriptors only to complete the results) in order to fasten the initial repetition detection stage. This allows us to process 6 months of TV in a few minutes.

6.4.3. Multimedia browsing

Participant: Laurent Amsaleg.

Traditionally, research in multimedia has focused primarily on analyzing and understanding the contents of media documents, by defining clever ways to extract relevant information from the multimedia files, thereby hoping to eventually bridge the semantic gap. We have observed that much of the research in multimedia is trying to *link* the information automatically extracted from the contents to create a meaningful user-experience. Most of the state-of-art solutions are very ad-hoc, and we believe that multimedia is lacking a powerful and flexible data model where multimedia data (ranging from entire documents to elements automatically extracted from the contents such as faces, scenes, objects, ...) can be appropriately represented as well as the relationships between data items. Instead, we propose a multi-dimensional model for media browsing, called ObjectCube, based on the multi-dimensional model commonly used in On-Line Analytical Processing (OLAP) applications. This model has been implemented in a prototype called ObjectCube, and its performance evaluated using personal photo collections of up to one million images. We also worked on exposing plug-in API for image analysis and browsing methods, facilitating the use of the prototype and its model as a demonstration platform.

6.4.4. Video summarization

Participants: Mohamed-Haykel Boukadida, Patrick Gros.

Joint work with Orange labs.

Up to now, most video summarization methods are based on concepts like saliency and often use a single modality. In order to develop a more general framework, we propose to use a constraint programming approach, where summarizing a video is seen as a constraint resolution problem, which consists in choosing certains excerpts with respect to various criteria. This first year of work on the topic was mainly devoted to discover the abilities of Choco, a constraint solver, and to study how summarization can be formulated as a constraint resolution problem.

6.4.5. Graph organization of large scale news archives

Participants: Guillaume Gravier, Ludivine Kuznik, Pascale Sébillot.

This work is done in collaboration with Jean Carrive at Institut National de l'Audiovisuel in the framework of a joint Ph. D. thesis within the Quaero project.

The idea of this work is to automatically create links and threads between reports in several years of broadcast news shows, based either on the documentary records of the shows and/or on the automatic transcripts. We studied how standard information retrieval measures of similarity can be used to build an epsilon-nearest neighbor graph from the various fields of the documentary records. Depending on the field used (title, keywords from a thesaurus, summary, speech transcript) and the metrics, different types of clusters can be obtained in the graph. We proposed metrics mimicking recall and precision on documents to analyze the graphs obtained and quantify the potential interest of various graph construction strategies for topic threading.

6.5. Language processing in multimedia

6.5.1. Lexical-phonetic automata for spoken utterance indexing and retrieval

Participants: Julien Fayolle, Guillaume Gravier, Fabienne Moreau, Christian Raymond.

This work was partly done in the context of the Quaero project.

Spoken content retrieval relies on the fields of automatic speech recognition and information retrieval (IR). However, IR tools made for text are not adapted to automatic transcripts which are particularly incomplete and uncertain. Even if in-vocabulary words are usually well-recognized, these transcripts contain many recognition errors affecting notably out-of-vocabulary words and named entities that convey important discourse information (e.g., person names, localizations, organizations) necessary for IR. This year, we have proposed a method for indexing spoken utterances which combines lexical and phonetic hypotheses in a hybrid index built from automata [35], [36]. The retrieval is performed by a lexical-phonetic and semi-imperfect matching whose aim is to improve the recall. A feature vector, containing edit distance scores and a confidence measure, weights each transition to help the filtering of the candidate utterance list for a more precise search. We have demonstrated the complementarity of the lexical and phonetic levels (extracted from the 1-best speech recognition hypothesis) and the advantage of using a hybrid index, a semi-imperfect matching and a supervised filtering (combining edit distance scores and a confidence measure).

6.5.2. *Information extraction and text mining*

Participants: Ali Reza Ebadat, Vincent Claveau, Pascale Sébillot.

This work was partly done in the framework of the Quaero project.

In the framework of Ali-Reza Ebadat's thesis on information extraction for multimedia analysis, we have investigated techniques for robust text-mining on texts or speech transcripts. We have developed several supervised models:

- entity detection and entity classification; the goal is to detect, into a text, pre-defined categories of entities and to label them accordingly. The techniques that we developed cascade chunk parsing with simple classification tools, resulting in a very efficient and simple to train NLP tool.
- relation detection; this model relies on k-NN approach with a language-modeling based distance. Since it relies on surface elements, it can handle noisy data such as speech transcripts.

We have also developed unsupervised models for information discovery:

- entity clustering; the goal is to detect and group, without a priori knowledge, entities. We have shown that weighting techniques used in information retrieval can be used as relevant features to describe the entity.
- relation clustering: as for entity, the goal is to group relations (that is, pairs of entities) without a priori or pre-defined categories. Our approach is pioneer is this field and relies on clustering with language-modeling based distances.

Some of these models have been evaluated in the framework of the Quaero evaluation campaign and TexMex ranked first in three of the tracks (entity detection and categorization) and close second in the last one (relation detection and categorization).

6.5.3. *Morphological analysis for information retrieval*

Participants: Vincent Claveau, Ewa Kijak.

In the biomedical field, the key to access information is the use of specialized terms (like *photochemotherapy*). These complex morphological structures may prevent a user querying for *gastrodynia* to retrieve texts containing *stomachalgia*. In that context, we have developed a new unsupervised technique to identify the various meaningful components of these terms and use this analysis to improve biomedical information retrieval. Our approach combines an automatic alignment using a pivot language, and an analogical learning that allows an accurate morphological analysis of terms. We ave shown that these morphological analyses can be used to greatly improve the indexing of medical documents.

6.5.4. *Unsupervised hierarchical topic segmentation*

Participants: Guillaume Gravier, Pascale Sébillot, Anca-Roxana Simon.

Linear topic segmentation has been widely studied for textual data and recently adapted to spoken contents. However, most documents exhibit a hierarchy of topics which cannot be recovered using linear segmentation. We investigated hierarchical topic segmentation of TV programs exploiting the spoken material. Recursively applying linear segmentation methods is one solution but fails at the lowest levels of the hierarchy when small segments are targeted, in particular when transcription errors jeopardize lexical cohesion. To skirt these issues, we investigated the use of indirect comparison between segments via vectorization techniques at the lower level of the hierarchy, using simple segmentation methods based on TextTiling. Results were similar to those obtained by the recursive use of a more elaborate probabilistic topic segmentation method. Future work will focus on using indirect comparison within the probabilistic framework.

6.6. Competitions and international evaluation campaign

6.6.1. *Mediaeval's affect task: Violent scenes detection task*

Participants: Guillaume Gravier, Patrick Gros, Cédric Penet.

The project-team participated in the Affect Task of the MediaEval 2012 benchmark, both as part of the organizing team and as competitor [64], [67].

6.6.2. Mediaeval's placing task: Geo-localization of videos

Participants: Jonathan Delhumeau, Guillaume Gravier, Hervé Jégou, Michele Trevisiol.

This work was partly done in the context of the Quaero project.

We developed an efficient and effective system to identify the geographic location of videos using a multimodal cascade of techniques exploiting all available sources of information, from user assigned tags to user data and image content. We also proposed a novel hierarchical strategy to exploit tags using information retrieval techniques. A coarse geographic area is first identified before refining the search to find exact geo-coordinates. Area and coordinates are obtained from a vector space representation of the tags using appropriate weighting and normalization [68].

We participated in the Placing Task of the MediaEval 2012 benchmark, where we ranked first on one of the mandatory runs (no gazeteers, no dictionary).

6.6.3. Mediaeval: Search & hyperlinking

Participants: Guillaume Gravier, Camille Guinaudeau, Pascale Sébillot.

We participated in the Search and Hyperlinking task proposed in the framework of the MediaEval benchmark initiative in 2012. We developed a solution for the hyperlinking subtask in which participants were required to return a ranked list of video segments potentially relevant to the answer provided for an initial query, thus creating links between video segments.

Our solution, based on information retrieval techniques, implements two separate module: The retrieval of relevant videos, followed by the selection of short segments specifically corresponding to the information need. First, the hyperlinking module computes the similarity between a video segment query and the collection of videos and returns a ranked list of relevant videos. We investigated several parameterization and ranking strategies. In the second step, we extract from each video the segment that is the closest, from a meaning point of view, to the video segment query, using topic segmentation methods [42].

Our system ranked either first or second depending on the evaluation conditions.

6.6.4. ETAPE named entities evaluation campaign

Participant: Christian Raymond.

Christian Raymond participated to the ETAPE Named Entities evaluation campaign. The goal was to propose a system able to tag NE following the new tree-stuctured NE definition given in the Quaero project. The evaluation has been done on manual and 5 automatic transcriptions of french TV and Radio shows produced by 5 different automatic speech recognition systems. The system was ranked first with results far better than those of the other participating systems.

6.6.5. DEFT evaluation campaign participation

Participants: Vincent Claveau, Christian Raymond.

Christian Raymond and Vincent Claveau participated to DEFT. The task proposed was to work on a corpus of scientific papers, by focusing the work on the issue of indexing the scientific papers: identifying the keywords chosen by the authors to index their paper, considering both abstract and whole article. Two tasks were proposed which led them to test two different strategies . For the first task, a list of keywords was provided. Based on that, our first strategy is to consider that as an Information Retrieval problem in which the keywords are the queries that are attributed to the best ranked documents. This approach yielded very good results. For the second task, only the articles were known. For this task, our approach is mainly based on a term extraction system whose results are reordered by a machine learning [27] technique.

6.6.6. Trecvid: Multimedia Indexing task

Participants: Jonathan Delhumeau, Philippe-Henri Gosselin, Hervé Jégou.

This work was partly done in the context of the Quaero project.

Texmex has taking part to the Quaero [50] and IRIM [21] submissions of Trecvid in the Multimedia indexing task, by providing some state-of-the-art image descriptors and collaborating with the LIG to set up the dimensionality reduction tool for high-dimensional vectors. The Quaero Rank was ranked 3rd in the full task (1st amongst European submissions).

VR4I Team

6. New Results

6.1. Physical modelling and simulation

6.1.1. Real-time mechanical simulation of brittle fracture

Participants: Loeïz Glondu, Georges Dumont [contact], Maud Marchal [contact].

Simulating brittle fracture of stiff bodies is now commonplace in computer graphics. However, simulating the deformations undergone by the bodies in a realistic way remains computationaly expensive. Thus, physically-based simulation of brittle fracture in real-time is still challenging for interactive applications. We have worked on a physically-based approach for simulating realistic brittle fracture in real-time.Our method is mainly composed of two parts: (1) a fracture initiation method based on modal analysis, (2) a fast energy-based fracture propagation algorithm. Results that emphasize the "real-time" part of this method have been published in [9]. Collision detection plays a key role in simulation performance. This is particularly true for fracture simulation, where multiple new objects are dynamically created. We proposed algorithms and data structures for collision detection in real-time brittle fracture simulations. We build on a combination of well-known efficient data structures, namely distance fields and sphere trees, making our algorithm easy to integrate on existing simulation engines. We proposed novel methods to construct these data structures, such that they can be efficiently updated upon fracture events and integrated in a simple yet effective self-adapting contact selection algorithm. Altogether, we drastically reduce the cost of both collision detection and collision response. We have evaluated our global solution for collision detection on challenging scenarios, achieving high frame rates suited for hard real-time applications such as video games or haptics [23]. Moreover, a common weathering effect is the appearance of cracks due to material fractures. We introduced a method to exemplar-based modeling that creates weathered patterns on synthetic objects by matching the statistics of fracture patterns in a photograph. A user study was proposed to determine which statistics are correlated to visual similarity and how they are perceived by the user. A revised physically-based fracture model capable of producing a wide range of crack patterns at interactive rates has been proposed whose parameter can be determined by a Bayesian optimization to produce a pattern with the same key statistics as an exemplar [10]. This work was the subject of the PhD thesis of Loeïz Glondu that has been successfuly defensed [3].

6.1.2. Collision detection in large scale environments with High Performance Computing

Participants: Bruno Arnaldi, Quentin Avril, Valérie Gouranton [contact].

We propose [14] a novel and efficient GPU-based parallel algorithm to cull non-colliding objects pairs in very large scale dynamic simulations. It allows to cull objects in less than 25ms with more than 100K objects. It is designed for many-core GPU and fully exploits multi-threaded capabilities and data-parallelism. In order to take advantage of the high number of cores, a new mapping function is defined that enables GPU threads to determine the objects pair to compute without any global memory access. These new optimized GPU kernel functions use the thread indexes and turn them into a unique pair of objects to test. A square root approximation technique is used based on Newton's estimation, enabling the threads to only perform a few atomic operations to cull non-colliding objects. We present a first characterization of the approximation errors that enables the fixing of incorrect computations. Input and output GPU streams are optimized using binary masks. The implementation and evaluation is made on large-scale dynamic rigid body simulations. The increase in speed is highlighted over other recently proposed CPU and GPU-based techniques. The comparison shows that our system is, in most cases, faster than previous approaches.

6.1.3. Simulation evaluations for ergonomics in VR

Participants: Georges Dumont [contact], Charles Pontonnier.

The use of virtual reality tools for ergonomics applications is a very important challenge.

In order to improve the design of workstations, an estimation of the muscle forces involved in the work tasks has to be done.

For example, one of our study assessed the level of confidence for results obtained with an inverse dynamics method from real captured work tasks. The chosen tasks are meat cutting tasks, well known to be highly correlated to musculoskeletal troubles appearance in the slaughter industry.

The experimental protocol consists in recording three main data during meat cutting tasks, and analyze their variation when some of the workstation design parameters are changing.

1. External (cutting)force data : for this purpose, a 3D instrumented knife has been designed in order to record the force applied by the subject during the task;
2. Motion Capture data : for this purpose, we use a motion capture system with active markers (Visualeyez II, Phoenix Technologies, Canada);
3. EMG data : several muscle activities are recorded using electromyographic electrodes, in order to compare these activities to the ones obtained from the inverse dynamics method.

With regard to the design parameters, that are the table height and the cutting direction, trends of recorded muscles activations were defined in order to be compared to computed ones issued from a musculoskeletal simulation performed with the AnyBody modeling system (AnyBody, Aalborg, Denmark). Results showed that an optimal set of design parameters can be obtained [27], whereas motor control strategies are highly dependent to the subject's experience and morphology.

This work has been done in collaboration with the Center for Sensory-motor Interaction (SMI, Aalborg University, Aalborg, Denmark), particularly Mark de Zee (Associate Professor) and Pascal Madeleine (Professor).

Furthermore, the fidelity of the VR simulator has to be evaluated (see Figure 2). For example, a simulator for assembly task has been evaluated in comparing different types of interaction : real, virtual and virtual + force feedback [28]. Objective and subjective metrics of discomfort led to highlight the influence of the environment on motor control and sensory feedback, changing more or less deeply the way the task is performed. Those change have to be taken into account to enable the use of such simulators for ergonomics purposes.

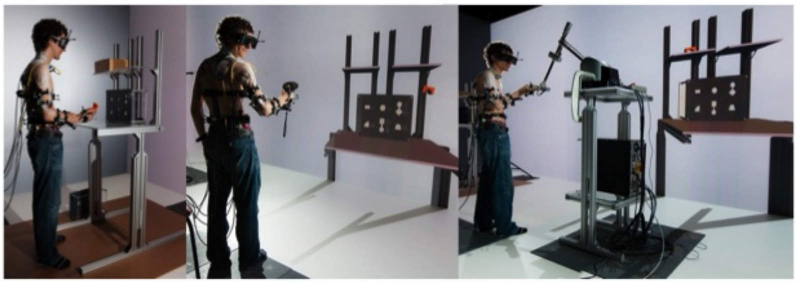


Figure 2. Simulation of an assembly task (Left in real, center in virtual, right in virtual with force-feedback)

6.2. Multimodal immersive interaction

6.2.1. Immersive Archaeology

Participants: Bruno Arnaldi, Georges Dumont, Ronan Gaugne [contact], Valérie Gouranton [contact].

We propose a workflow of tools and procedures to reconstruct an existing archaeological site as a virtual 3D reconstitution in a large scale immersive system [35] . This interdisciplinary endeavor, gathering archaeologists and virtual reality computer scientists, is the first step of a joint research project with three objectives: (i) propose a common workflow to reconstruct archaeological sites as 3D models in fully immersive systems, (ii) provide archaeologists with tools and interaction metaphors to exploit immersive reconstitutions, and (iii) develop the use and access of immersive systems to archaeologists. In this context, we present [21] results from the immersive reconstitution of Carn's monument central chamber, in Finistere, France, a site currently studied by the Creaah archaeology laboratory. The results rely on a detailed workflow we propose, which uses efficient solutions to enable archaeologists to work with immersive systems. In particular, we proposed a procedure to model the central chamber of the Carn monument, and compare several softwares to deploy it in an immersive structure. We then proposed two immersive implementations of the central chamber, with simple interaction tools.

6.2.2. Novel 3D displays and user interfaces

Participants: Anatole Lécuyer [contact], David Gomez, Fernando Argelaguet, Maud Marchal, Jerome Ardouin.

We describe hereafter our recent results in the field of novel 3D User Interfaces and, more specifically, novel displays and interactive techniques to better perceive and interact in 3D. This encloses: (1) Novel interactive techniques for interaction with 3D web content, and (2) A novel display for augmented 3D vision.

6.2.2.1. Novel interactive techniques for 3D web content

The selection and manipulation of 3D content in desktop virtual environments is commonly achieved with 2D mouse cursor-based interaction. However, by interacting with image-based techniques we introduce a conflict between the 2D space in which the 2D cursor lays and the 3D content. For example, the 2D mouse cursor does not provide any information about the depth of the selected objects. In this situation, the user has to rely on the depth cues provided by the virtual environment, such as perspective deformation, shading and shadows.

In [24], we have explored new metaphors to improve the depth perception when interacting with 3D content. Our approach focus on the usage of 3D cursors controlled with 2D input devices (the Hand Avatar and the Torch) and a pseudo-motion parallax effect. The additional depth cues provided by the visual feedback of the 3D cursors and the motion parallax are expected to increase the users' depth perception of the environment.

The evaluation of proposed techniques showed that users depth perception was significantly increased. Users were able to better judge the depth ordering of virtual environment. Although 3D cursors showed a decrease of selection performance, it is compensated by the increased depth perception.

6.2.2.2. FLyVIZ : A novel display for providing humans with panoramic vision

Have you ever dreamed of having eyes in the back of your head? In [12], we have presented a novel display device called FlyVIZ which enables humans to experience a real-time 360-degree vision of their surroundings for the first time.

To do so, we combined a panoramic image acquisition system (positioned on top of the user's head) with a Head-Mounted Display (HMD). The omnidirectional images are transformed to fit the characteristics of HMD screens. As a result, the user can see his/her surroundings, in real-time, with 360 degree images mapped into the HMD field of view.

We foresee potential applications in different fields where augmented human capacity (an extended field-of-view) could benefit, such as surveillance, security, or entertainment. FlyVIZ could also be used in novel perception and neuroscience studies.

6.2.3. Brain-Computer Interfaces

Participants: Anatole Lécuyer [contact], Laurent George, Laurent Bonnet, Jozef Legeny.

Brain-computer interfaces (BCI) are communication systems that enable to send commands to a computer using only the brain activity. Cerebral activity is generally sensed with electroencephalography (or EEG). We describe hereafter our recent results in the field of brain-computer interfaces and virtual environments: (1) Novel signal processing techniques for EEG-based Brain-Computer Interfaces, and (2) Design and study of Brain-Computer Interaction with real and virtual environments.

6.2.3.1. Novel signal processing techniques for EEG-based Brain-Computer Interfaces

A first part of the BCI research conducted in the team is dedicated to EEG signal processing and classification techniques applied to cerebral EEG data.

To properly and efficiently decode brain signals into computer commands the application of efficient machine-learning techniques is required.

In [5] we could introduce two new features for the design of electroencephalography (EEG) based Brain-Computer Interfaces (BCI): one feature based on multifractal cumulants, and one feature based on the predictive complexity of the EEG time series. The multifractal cumulants feature measures the signal regularity, while the predictive complexity measures the difficulty to predict the future of the signal based on its past, hence a degree of how complex it is. We have conducted an evaluation of the performance of these two novel features on EEG data corresponding to motor-imagery. We also compared them to the gold standard features used in the BCI field, namely the Band-Power features. We evaluated these three kinds of features and their combinations on EEG signals from 13 subjects. Results obtained show that our novel features can lead to BCI designs with improved classification performance, notably when using and combining the three kinds of feature (band-power, multifractal cumulants, predictive complexity) together.
Evolutionary algorithms have also been increasingly applied in different steps of BCI implementations. In [29], we could then introduce the use of the covariance matrix adaptation evolution strategy (CMA-ES) for BCI systems based on motor imagery. The optimization algorithm was used to evolve linear classifiers able to outperform other traditional classifiers. We could also analyze the role of modeling variables interactions for additional insight in the understanding of the BCI paradigms.

6.2.3.2. Brain-Computer Interaction with real and virtual environments

A second part of our BCI research is dedicated to the improvement of BCI-based interaction with real and virtual environments. We have first initiated research on **Combining Haptic and Brain-Computer Interfaces**.

In [22], we have introduced the combined use of Brain-Computer Interfaces (BCI) and Haptic interfaces. We proposed to adapt haptic guides based on the mental activity measured by a BCI system. This novel approach has been illustrated within a proof-of-concept system: haptic guides were toggled during a path-following task thanks to a mental workload index provided by a BCI. The aim of this system was to provide haptic assistance only when the user's brain activity reflects a high mental workload.

A user study conducted with 8 participants showed that our proof-of-concept is operational and exploitable. Results showed that activation of haptic guides occurs in the most difficult part of the path-following task. Moreover it allowed to increase task performance by activating assistance only 59 percents of the time. Taken together, these results suggest that BCI could be used to determine when the user needs assistance during haptic interaction and to enable haptic guides accordingly.

This work paves the way to novel passive BCI applications such as medical training simulators based on passive BCI and smart guides. It has received the Best Paper Award of Eurohaptics 2012 conference, and was nominated for the BCI Award 2012.

6.2.4. Natural Interactive Walking in Virtual Environments

Participants: Anatole Lécuyer [contact], Maud Marchal [contact], Gabriel Cirio, Tony Regia Corte, Sébastien Hillaire, Léo Terziman.

Figure 3. Proof-of-concept system combining Haptic and a Brain-Computer Interface (haptic guides are toggled based on a mental workload index computed by the BCI)

We describe hereafter our recent results obtained in the field of "augmented" or "natural interactive" walking in virtual environments. Our first objective is to better understand the properties of human perception and human locomotion when walking in virtual worlds. Then, we intend to design advanced interactive techniques and interaction metaphors to enhance, in a general manner, the navigation possibilities in VR systems. Last, our intention is to improve the multisensory rendering of human locomotion and human walk in virtual environments, making full use of both haptic and visual feedback.

6.2.4.1. Perception of ground affordances in virtual environments

We have evaluated the perception of **ground affordances in virtual environments** (VE).

In [11], we considered the affordances for standing on a virtual slanted surface. Participants were asked to judge whether a virtual slanted surface supported upright stance. The objective was to evaluate if this perception was possible in virtual reality (VR) and comparable to previous works conducted in real environments. We found that the perception of affordances for standing on a slanted surface in virtual reality is possible and comparable (with an underestimation) to previous studies conducted in real environments. We also found that participants were able to extract and to use virtual information about friction in order to judge whether a slanted surface supported an upright stance. Finally, results revealed that the person's position on the slanted surface is involved in the perception of affordances for standing on virtual grounds. Taken together, our results show quantitatively that the perception of affordances can be effective in virtual environments, and influenced by both environmental and person properties. Such a perceptual evaluation of affordances in VR could guide VE designers to improve their designs and to better understand the effect of these designs on VE users.

6.2.4.2. Novel metaphors for navigating virtual environments

Immersive spaces such as 4-sided displays with stereo viewing and high-quality tracking provide a very engaging and realistic virtual experience. However, walking is inherently limited by the restricted physical space, both due to the screens (limited translation) and the missing back screen (limited rotation).

In [7], we proposed three novel locomotion techniques that have three concurrent goals: keep the user safe from reaching the translational and rotational boundaries; increase the amount of real walking and finally, provide a more enjoyable and ecological interaction paradigm compared to traditional controller-based approaches.

We notably introduced the "Virtual Companion", which uses a small bird to guide the user through VEs larger than the physical space. We evaluated the three new techniques through a user study with travel-to-target and path following tasks. The study provided insight into the relative strengths of each new technique for the three aforementioned goals. Specifically, if speed and accuracy are paramount, traditional controller interfaces augmented with our novel warning techniques may be more appropriate; if physical walking is more important, two of our paradigms (extended Magic Barrier Tape and Constrained Wand) should be preferred; last, fun and ecological criteria would favor the Virtual Companion.

6.2.4.3. Novel sensory feedback for improving sensation of walking in VR: the King-Kong Effects

Third, we have designed novel sensory feedbacks named "King-Kong Effects" to enhance the sensation of walking in virtual environments [33].

King Kong Effects are inspired by special effects in movies in which the incoming of a gigantic creature is suggested by adding visual vibrations/pulses to the camera at each of its steps (Figure 4).

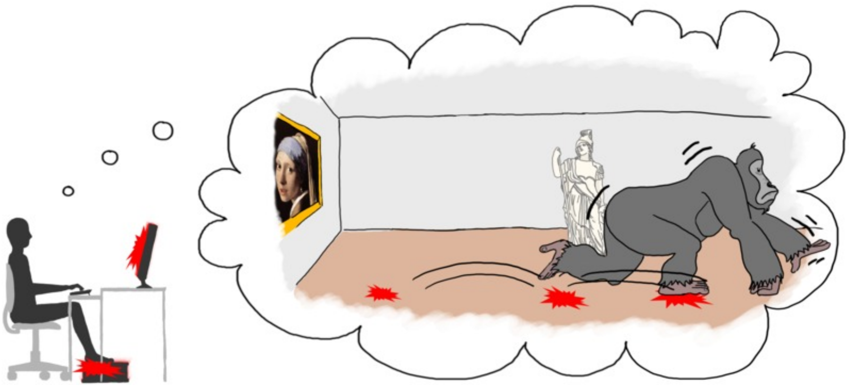


Figure 4. Concept of the King Kong Effects: Visual and Tactile vibrations inspired by special effects in movies enhance the sensation of walking in VE. Visual and Tactile feedbacks are generated at each step made in the VE.

We thus proposed to add artificial visual or tactile vibrations (King-Kong Effects or KKE) at each footstep detected (or simulated) during the virtual walk of the user. The user can be seated, and our system proposes to use vibrotactile tiles located under his/her feet for tactile rendering, in addition to the visual display. We have designed different kinds of KKE based on vertical or lateral oscillations, physical or metaphorical patterns, and one or two peaks for heal-toe contacts simulation.

We have conducted different experiments to evaluate the preferences of users navigating with or without the various KKE. Taken together, our results identify the best choices in term of sensation of walking for future uses of visual and tactile KKE, and they suggest a preference for multisensory combinations. Our King-Kong effects could be used in a variety of VR applications targeting the immersion of a user walking in a 3D virtual scene.

6.2.5. Haptic Interaction

Participants: Fernando Argelaguet, Fabien Danieau, Anatole Lécuyer [contact], Maud Marchal, Anthony Talvas.

6.2.5.1. Pseudo-Haptic Feedback

Pseudo-haptic feedback is a technique meant to simulate haptic sensations in virtual environments using visual feedback and properties of human visuo-haptic perception. Pseudo-haptic feedback uses vision to distort haptic perception and verges on haptic illusions. Pseudo-haptic feedback has been used to simulate various haptic properties such as the stiffness of a virtual spring, the texture of an image, or the mass of a virtual object.

In [13], we focused on the improvement of pseudo-haptic textures. Pseudo-haptic textures allow to optically-induce relief in tex- tures without a haptic device by adjusting the speed of the mouse pointer according to the depth information encoded in the texture. In this work, we have presented a novel approach for using curvature information instead of relying on depth information. The curvature of the texture is encoded in a normal map which allows the computation of the curvature and local changes of orientation, according to the mouse position and direction.

A user evaluation was conducted to compare the optically-induced haptic feedback of the curvature-based approach versus the original depth-based approach based on depth maps. Results showed that users, in addition to being able to efficiently recognize simulated bumps and holes with the curvature-based approach, were also able to discriminate shapes with lower frequency and amplitude.

6.2.5.2. Bi-Manual Haptic Feedback

In the field of haptics and virtual reality, two-handed interaction with virtual environments (VEs) is a domain that is slowly emerging while bearing very promising applications.

In [32] we could present a set of novel interactive techniques adapted to two-handed manipulation of objects with dual 3DoF single- point haptic devices (see Figure 5). We first proposed the double bubble for bimanual haptic exploration of virtual environments through hybrid position/rate controls, and a bimanual viewport adaptation method that keeps both proxies on screen in large environments. We also presented two bimanual haptic manipulation techniques that facilitate pick-and-place tasks: the joint control, which forces common control modes and control/display ratios for two interfaces grabbing an object, and the magnetic pinch, which simulates a magnet-like attraction between both hands to prevent unwanted drops of that object.

An experiment was conducted to assess the efficiency of these techniques for pick-and-place tasks, by comparing the double bubble with viewport adaptation to the clutching technique for extending the workspaces, and by measuring the benefits of the joint control and magnetic pinch.

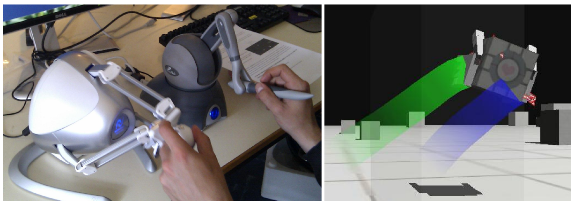


Figure 5. Bimanual pick-and-place task in a large virtual environment. A bimanual haptic setup made of two single-point devices (on left) allows to carry and displace a virtual cube using novel interactive techniques (on right).

6.2.5.3. Haptic Feedback and Haptic Seat for Enhancing AudioVisual Experience

This work aims at enhancing a classical video viewing experience by introducing realistic haptic feelings in a consumer environment.

First, in [16] a complete framework to both produce and render the motion embedded in an audiovisual content was proposed to enhance a natural movie viewing session. We especially considered the case of a first-person point of view audiovisual content and we propose a general workflow to address this problem. This latter includes a novel approach to both capture the motion and video of the scene of interest, together with a haptic rendering system for generating a sensation of motion. A complete methodology to evaluate the relevance of our framework was finally proposed and could demonstrate the interest of our approach.

Second, leveraging on the techniques and framework introduced previously, in [17] we could introduce a novel way of simulating motion sensations without calling for expensive and cumbersome motion plat- forms. The main idea consists in applying multiple force- feedbacks on the user's body to generate a sensation of motion while seated and experiencing passive navigation. A set of force-feedback devices are therefore arranged around a seat, as if various components of the seat could apply forces on the user, like mobile armrests or headrest. This new approach is called HapSeat (see Figure 6). A proof-of-concept has been designed within a structure which relies on 3 low-cost force-feedback devices, and two models were implemented to control them.

Results of a first user study suggests that subjective sensations of motion can be generated by both approaches. Taken together, our results pave the way to novel setups and motion effects for consumer living-places based on the HapSeat.

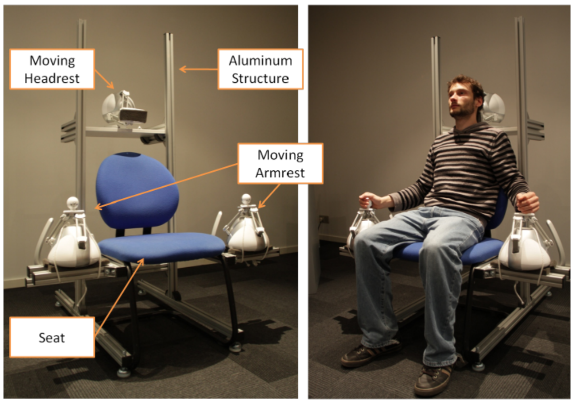


Figure 6. Prototype of the HapSeat. Left: seat structure with 3 force-feedback devices. Right: the system in use.

6.2.6. Interactions within 3D virtual universes

Participants: Thierry Duval [contact], Thi Thuong Huyen Nguyen, Cédric Fleury.

We have proposed some new metaphors allowing a guiding user to be fully aware of what the main user was seeing in the virtual universe and of what were the physical constraints of this user. We made a first prototype that made it possible to participate to the 3DUI 2012 contest [26], then we made further experiments showing the interest of the approach, these results will be presented in [25].

Our work focuses upon new formalisms for 3D interactions in virtual environments, to define what an interactive object is, what an interaction tool is, and how these two kinds of objects can communicate together. We also propose virtual reality patterns to combine navigation with interaction in immersive virtual environments. We are currently working about new multi-point interaction techniques to allow users to precisely manipulate virtuel objects.

6.3. Collaborative work in CVE's

6.3.1. The immersive interactive virtual cabin (IIVC)

Participants: Thierry Duval [contact], Valérie Gouranton [contact], Alain Chauffaut, Bruno Arnaldi, Cédric Fleury, Thi Thuong Huyen Nguyen, Georges Dumont.

We are still improving the architecture of our Immersive Interactive Virtual Cabin to improve the user's immersion with all his real tools and so to make the design and the use of 3D interaction techniques easier, and to make possible to use them in various contexts, either for different kinds of applications, or with different kinds of physical input devices.

The IIVC is now fully implemented in our two VR platforms: OpenMASK 5.1 and Collaviz 7.1.2 .

We have used the IIVC in order to provide efficient collaboration between users in a guiding task, allowing a guiding user to be fully aware of what the main user was seeing in the virtual universe and of what were the physical constraints of this user. We made a first prototype that made it possible to participate to the 3DUI 2012 contest [26], then we made further experiments showing the interest of the approach, these results will be presented in [25]. We also proposed to use the IIVC to enhance the communication between users sharing a virtual universe by helping them to build a cognitive model of the other users' environment [19]

6.3.2. Generic architecture for 3D interoperability

Participants: Thierry Duval [contact], Valérie Gouranton, Cédric Fleury, Rozenn Bouville Berthelot, Bruno Arnaldi.

Our goal is to allow software developers to build 3D interactive and collaborative environments without bothering with the 3D graphics API they are using. This work is the achievement of the IIVC software architecture. We have proposed PAC-C3D (Figure 7), a new software architectural model for collaborative 3D applications, in order to provide a higher abstraction for designing 3D virtual objects, and in order to provide interoperability, making it possible to share a virtual universe between heterogeneous 3D viewers.

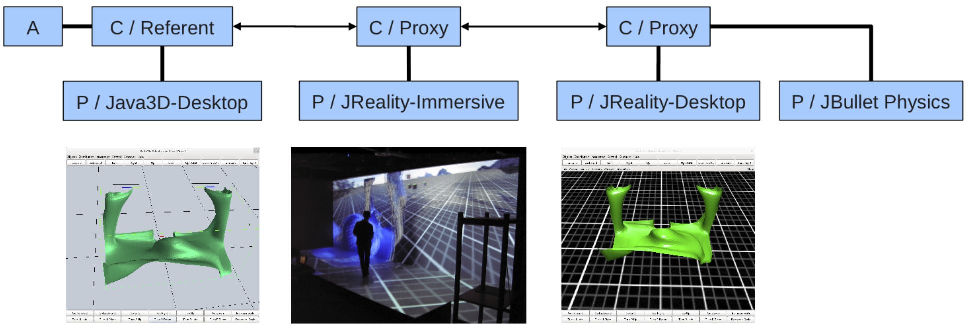


Figure 7. The PAC-C3D software architectural model makes interoperability possible between heterogeneous 3D viewers

We also study how to offer interoperability between virtual objects that are loaded in the same virtual environment but that are described using different formats. This is why we have proposed a generic architecture for enabling interoperability between 3D formats (Figure 8), the Scene Graph Adapter. Our SGA is now able to allow events coming from a 3D format to act upon data provided in another format, such as X3D events operating on Collada data, and makes also it possible to compose different format files [15].

6.3.3. Collaborative interaction model

Participants: Bruno Arnaldi, Valérie Gouranton [contact], Andrés Saraos Luna.

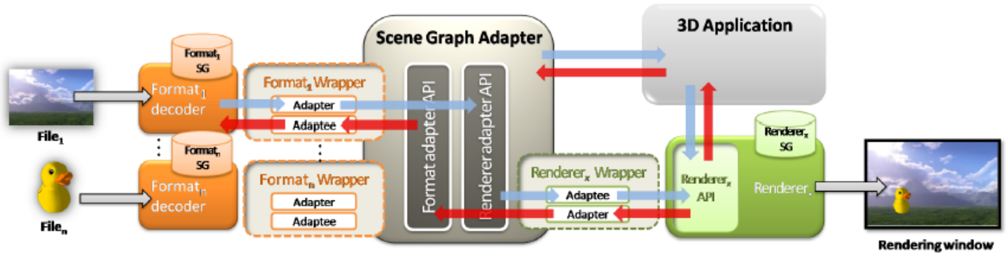


Figure 8. Our architecture allows the loading of any 3D graphics format simultaneously in any available rendering engine. The scene graph adapter is an interface that adapts a scene graph (SG) of a given format into a renderer scene graph and which also allows the rendering part to request this scene graph.

Our work ponders on collaborative interactions in Collaborative Virtual Environments for Training, with an emphasis on collaborative interactions between Real Humans and Virtual Humans working as a team. We propose [30] a new collaborative interaction model and from it construct a set of tools to describe and define such collaborative interactions [34].

6.4. Immersia Virtual Reality room

Participants: Georges Dumont [contact], Alain Chauffaut, Ronan Gaugne [contact], Marwan Badawi.

The team was the first in France to host a large-scale immersive virtual reality equipment known as Immersia (see figure 9). This platform, with full visual and sound immersion, is dedicated to real-time, multimodal (vision, sound, haptic, BCI) and immersive interaction. It will accommodate experiments using interactive and collaborative virtual-reality applications that have multiple local or remote users. Our new wall has four faces: a front, two sides and a ground. Dimensions are 9.6 m wide, 2.9 m deep and 3.1 m hight. The visual reproduction system combines ten Barco Galaxy NW12 projectors and three Barco Galaxy 7+ projectors. Visual images from Barco projectors are rendered on glass screens. They are adjusted for the users position, and this, together with their high resolution and homogeneous colouring, make them very realistic. The ART localization system, constituted of 16 ARTtrack 2 cameras, enables real objects to be located within the U-shape. Sound rendering is provided by a Yamaha processor, linked either to Genelec speakers with 10.2 format sound or Beyer Dynamic headsets with 5.1 virtual format sound, controlled by the users position.

The Immersia Virtual Reality room has been inaugurated on 2012, june, the 20th. We have hosted the project VR-GO, a Trans National Acces VISIONAIR project in june 2012. The goal was to evaluate an assembly by comparing different types of interaction : real, virtual and virtual + force feedback [28].

Figure 9. Immersia Virtual Reality Room

WAM Project-Team

6. New Results

6.1. Multimedia Models and Formats

In the context of the CLAIRE project (see section 7.1.1), a new model for educational documents has been defined. The objectives of this model are:

- to seamlessly handle conventional and richmedia content in the context of a unique pedagogical web platform.
- to be able to store and recover any multimedia document including its spatial and time structure, consistent with HTML5 and Timesheets specifications.
- to have a data model which is format agnostic to cope with existing and future rendering systems.
- to cope with the authoring needs of all users.

We have more specifically worked on the multimedia modelling part for defining spatial and temporal fragment types. These types are used to express the synchronization beween different elements within the document.

We are now using this model in the definition and implementation of a web-based authoring user interface.

6.2. XML Processing

In the area of XML processing, we have obtained new results in several directions:

- We have introduced the first system capable of statically verifying properties of a given cascading style sheet (CSS) over the whole set of documents to which this style sheet applies [5]. Properties include coverage of styling information and absence of erroneous rendering.
- In a joint work with the EXMO team, we have introduced a novel approach for deciding the SPARQL query containment problem in the presence of schemas, that paves the way for future extensions [4] [3] [8] [1].
- We have revisited the problem of XML Query-Update Independence Analysis, and showed the relevance of an approach that has been neglected in the literature so far [6]. In particular, we have compared an SMT-modulo with a tree logic approach to Independence Analysis.
- We have made progress on the characterization of the impacts of schema changes on XQuery programs [7].
- We have formally proved a result about the factorization power of the Lean: a construction that we use to speed up the XML Reasoning Solver. We have characterized which kind of duplicate subformulas this construction eliminates, and how [10].
- We have proposed a novel technique and a tool for the static type-checking of XQuery programs, using backward type inference [11].
- We have defined a type system for integrating session types for objects in object-oriented languages such as Java, with full structural subtyping, without altering the language semantics [9]. Session types are protocol specifications which describe which sequences of method calls are allowed or disallowed on a given object.

We briefly review these results below.

6.2.1. Automated Analysis of Cascading Style Sheets (CSS)

Developing and maintaining cascading style sheets (CSS) is an important issue to web developers as they suffer from the lack of rigorous methods. Most existing means rely on validators that check syntactic rules, and on runtime debuggers that check the behavior of a CSS style sheet on a particular document instance. However, the aim of most style sheets is to be applied to an entire set of documents, usually defined by some schema. To this end, a CSS style sheet is usually written w.r.t. a given schema. While usual debugging tools help reducing the number of bugs, they do not ultimately allow to prove properties over the whole set of documents to which the style sheet is intended to be applied. We have developed a novel approach to fill this lack [5]. The main ideas are borrowed from the fields of logic and compile-time verification and applied to the analysis of CSS style sheets. We have implemented an original tool (see section 5.1.1) based on recent advances in tree logics. The tool is capable of statically detecting a wide range of errors (such as empty CSS selectors and semantically equivalent selectors), as well as proving properties related to sets of documents (such as coverage of styling information), in the presence or absence of schema information. This new tool can be used in addition to existing runtime debuggers to ensure a higher level of quality of CSS style sheets.

6.2.2. Deciding Satisfiability and Containment for Semantic Web Queries

The problem of SPARQL query containment is defined as determining if the result of one query is included in the result of another for any RDF graph. Query containment is important in many areas, including information integration, query optimization, and reasoning about Entity-Relationship diagrams [1].

We encode this problem into an expressive logic called μ-calculus: where RDF graphs become transition systems, queries and schema axioms become formulas [4] [3]. Thus, the containment problem is reduced to formula satisfiability test. Beyond the logic's expressive power, satisfiability solvers are available for it. Hence, this study allows to exploit these advantages.

In addition, in order to experimentally assess implementation limitations, we have designed a benchmark suite offering different experimental settings depending on the type of queries, projection and reasoning (RDFS) [8]. We have applied this benchmark to three available systems using different techniques highlighting the strengths and weaknesses of such systems.

6.2.3. XML Query-Update Independence Analysis Revisited

XML transformations can be resource-costly in particular when applied to very large XML documents and document sets. Those transformations usually involve lots of XPath queries and may not need to be entirely re-executed following an update of the input document. In this context, a given query is said to be independent of a given update if, for any XML document, the results of the query are not affected by the update. We have revisited Benedikt and Cheney's framework for query-update independence analysis and we have shown that performance can be drastically enhanced, contradicting their initial claims [6]. The essence of our approach and results resides in the use of an appropriate logic, to which queries and updates are both succinctly translated. Compared to previous approaches, ours is more expressive from a theoretical point of view, equally accurate, and more efficient in practice. We have illustrated this through practical experiments and comparative figures.

6.2.4. Toward Automated Schema-directed Code Revision

Updating XQuery programs in accordance with a change of the input XML schema is known to be a time-consuming and error-prone task. We have designed an automatic method aimed at helping developers realign the XQuery program with the new schema [7]. First, we have devised a taxonomy of possible problems induced by a schema change. This allows to differentiate problems according to their severity levels, e.g. errors that require code revision, and semantic changes that should be brought to the developer's attention. Second, we have provided the necessary algorithms to detect such problems using our solver (see section 5.1) to check satisfiability of XPath expressions.

6.2.5. Logical Combinators for Rich Type Systems

We have developed a functional approach to design rich type systems based on an elegant logical representation of types [10]. The representation is not only clean but it also avoids exponential increases in combined complexity due to subformula duplication. This opens the way to solving a wide range of problems such as subtyping in exponential-time even though their direct translation into the underlying logic results in an exponential blowup of the formula size, yielding an incorrectly presumed two-exponential time complexity.

6.2.6. Backward type inference for XQuery

We have designed a novel technique and a tool for static type-checking of XQuery programs [11]. The tool looks for errors in the program by jointly analyzing the source code of the program, input and output schemas that respectively describe the sets of documents admissible as input and as output of the program. The crux and the novelty of our results reside in the joint use of backward type inference and a two-way logic to represent inferred tree type portions. This allowed us to design and implement a type-checker for XQuery which is more precise and supports a larger fragment of XQuery compared to the approaches previously proposed in the literature; in particular compared to the only few actually implemented static type-checkers such as the one in Galax. The whole system uses compilers and a satisfiability solver for deciding containment for two-way regular tree expressions. Our tool takes an XQuery program and two schemas S_{in} and S_{out} as input. If the program is found incorrect, then it automatically generates a counter-example valid w.r.t. S_{in} and such that the program produces an invalid output w.r.t S_{out}. This counter-example can be used by the programmer to fix the program.

6.2.7. Session types

Session types allow communication protocols to be specified type-theoretically so that protocol implementations can be verified by static type checking. In [9], we extend previous work on session types for distributed object-oriented languages in three ways. (1) We attach a session type to a class definition, to specify the possible sequences of method calls. (2) We allow a session type (protocol) implementation to be *modularized*, i.e. partitioned into separately-callable methods. (3) We treat session-typed communication channels as objects, integrating their session types with the session types of classes. The result is an elegant unification of communication channels and their session types, distributed object-oriented programming, and a form of typestate supporting non-uniform objects, i.e. objects that dynamically change the set of available methods. We define syntax, operational semantics, a sound type system, and a sound and complete type checking algorithm for a small distributed class-based object-oriented language with structural subtyping. Static typing guarantees that both sequences of messages on channels, and sequences of method calls on objects, conform to type-theoretic specifications, thus ensuring type-safety. The language includes expected features of session types, such as delegation, and expected features of object-oriented programming, such as encapsulation of local state. The main ideas have been implemented as a prototype, extending Java 1.4.

6.3. Multimedia Authoring

In cooperation with EPFL (Lausanne) we pursue our research on template-driven editing for XML multimedia contents (see section 3.3.2). Experiments with very different types of contents have been done with the AXEL library. AXEL is developed by EPFL, based on our joint work on template languages. It is an innovative multi-purpose client-side authoring framework intended for web users with limited skills.

We have addressed the issue of authoring XML multimedia content on the web, focusing on methods that apply to such contents as structured documents, factual data, and multimedia objects [2]. We have shown that a template-based approach enhances the ability for multiple applications to use the produced content.

6.4. Augmented Environments

Most results in the area of augmented environments were presented through various software products and prototypes, including:

- IXE, Interactive eXtensible Engine (see section 5.5 for details). In particular, IXE allowed us to show that a precision of one step is attainable, guidance being done through a mix of spatialized vocal instructions and 3D audio.
- GIF Demonstrator: This application was used to showcase our technologies at the Grenoble Innovation Fair (GIF). Augmented reality was used to find the various booths and products, while our indoor navigation system was guiding visitors to any booth.
- Interactive Audio Panorama: A fun interactive experience with virtual audio. It immerses the user in a complete 360° audio panorama and allows her/him to discover a futuristic house. It demonstrates the authoring possibilities offered by the MAUDL interactive audio language.
- PDRTrack: An indoor localization utility demonstrating the various correction parameters of our IMU-based localization system. The user can record data sets and simulate using various parameters to find out the effect of different map matching settings and their result on localization accuracy. The user can also simply walk in real-time with tracking enabled on a given OpenStreetMap network.
- Sugimotocho Stn: A model of this railway station has been built with the help of the GISLab (Osaka City University). An electronic kick-scooter was used to measure distances and a navigation network was designed to help people to move around in the station.

These products and prototypes were presented in various fora in 2012, in particular at:

- Grenoble Innovation Fair
- 4I Forum
- 6th European eAccessibility Forum
- State Of The Map 2012

WILLOW Project-Team

6. New Results

6.1. 3D object and scene modeling, analysis, and retrieval

6.1.1. People Watching: Human Actions as a Cue for Single View Geometry

Participants: Vincent Delaitre, Ivan Laptev, Josef Sivic, Alexei Efros [CMU], David Fouhey [CMU], Abhinav Gupta [CMU].

We present an approach which exploits the coupling between human actions and scene geometry. We investigate the use of human pose as a cue for single-view 3D scene understanding. Our method builds upon recent advances in still-image pose estimation to extract functional and geometric constraints about the scene. These constraints are then used to improve state-of-the-art single-view 3D scene understanding approaches. The proposed method is validated on a collection of monocular time lapse sequences collected from YouTube and a dataset of still images of indoor scenes. We demonstrate that observing people performing different actions can significantly improve estimates of 3D scene geometry.

This work has been published in [11].

6.1.2. Learning and Calibrating Per-Location Classifiers for Visual Place Recognition

Participants: Petr Gronát, Josef Sivic, Guillaume Obozinski [Inria SIERRA], Tomáš Pajdla [CTU in Prague].

The aim of this work is to localize a query photograph by finding other images depicting the same place in a large geotagged image database. This is a challenging task due to changes in viewpoint, imaging conditions and the large size of the image database. The contribution of this work is two-fold. First, we cast the place recognition problem as a classification task and use the available geotags to train a classifier for each location in the database in a similar manner to per-exemplar SVMs in object recognition. Second, as only few positive training examples are available for each location, we propose a new approach to calibrate all the per-location SVM classifiers using *only* the negative examples. The calibration we propose relies on a significance measure essentially equivalent to the p-values classically used in statistical hypothesis testing. Experiments are performed on a database of 25,000 geotagged street view images of Pittsburgh and demonstrate improved place recognition accuracy of the proposed approach over the previous work. The problem addressed in this work is illustrated in Figure 1 .

This work has been submitted to CVPR 2013.

6.1.3. What Makes Paris Look like Paris?

Participants: Josef Sivic, Carl Doersch [CMU], Saurabh Singh [UIUC], Abhinav Gupta [CMU], Alexei Efros [CMU].

Given a large repository of geotagged imagery, we seek to automatically find visual elements, e.g. windows, balconies, and street signs, that are most distinctive for a certain geo-spatial area, for example the city of Paris. This is a tremendously difficult task as the visual features distinguishing architectural elements of different places can be very subtle. In addition, we face a hard search problem: given all possible patches in all images, which of them are both frequently occurring and geographically informative? To address these issues, we propose to use a discriminative clustering approach able to take into account the weak geographic supervision. We show that geographically representative image elements can be discovered automatically from Google Street View imagery in a discriminative manner. We demonstrate that these elements are visually interpretable and perceptually geo-informative. The discovered visual elements can also support a variety of computational geography tasks, such as mapping architectural correspondences and influences within and across cities, finding representative elements at different geo-spatial scales, and geographically-informed image retrieval. Example result is shown in Figure 2 .

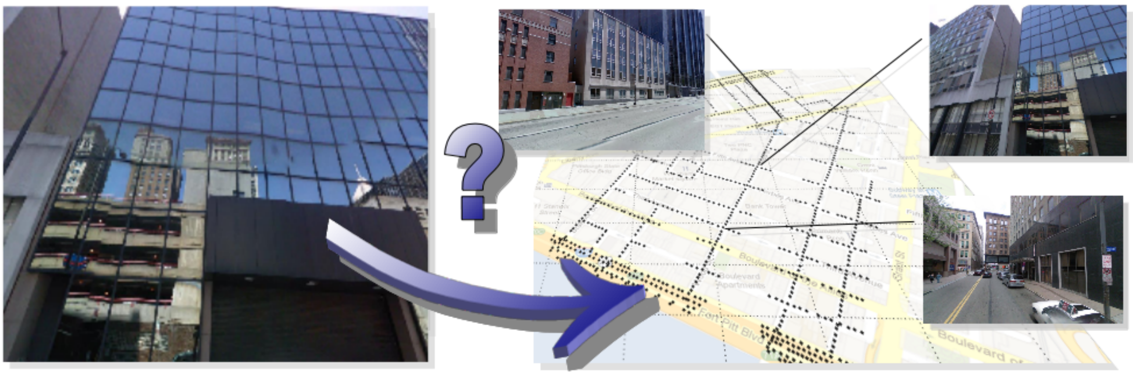


Figure 1. The goal of this work is to localize a query photograph (left) by finding other images of the same place in a large geotagged image database (right). We cast the problem as a classification task and learn a classifier for each location in the database. We develop a non-parametric procedure to calibrate the outputs of the large number of per-location classifiers without the need for additional positive training data.

Figure 2. Examples of geographic patterns in Paris (shown as red dots on the maps) for three discovered visual elements (shown below each map). Balconies with cast-iron railings are concentrated on the main boulevards (left). Windows with railings mostly occur on smaller streets (middle). Arch supporting columns are concentrated on Place des Vosges and the St. Germain market (right).

This work has been published in [6].

6.2. Category-level object and scene recognition

6.2.1. Task-Driven Dictionary Learning

Participants: Jean Ponce, Julien Mairal [Inria LEAR], Francis Bach [Inria SIERRA].

Modeling data with linear combinations of a few elements from a learned dictionary has been the focus of much recent research in machine learning, neuroscience and signal processing. For signals such as natural images that admit such sparse representations, it is now well established that these models are well suited to restoration tasks. In this context, learning the dictionary amounts to solving a large-scale matrix factorization problem, which can be done efficiently with classical optimization tools. The same approach has also been used for learning features from data for other purposes, e.g., image classification, but tuning the dictionary in a supervised way for these tasks has proven to be more difficult. In this paper, we present a general formulation for supervised dictionary learning adapted to a wide variety of tasks, and present an efficient algorithm for solving the corresponding optimization problem. Experiments on handwritten digit classification, digital art identification, nonlinear inverse image problems, and compressed sensing demonstrate that our approach is effective in large-scale settings, and is well suited to supervised and semi-supervised classification, as well as regression tasks for data that admit sparse representations.

This work has been published in [7].

6.2.2. Object Detection Using Strongly-Supervised Deformable Part Models

Participants: Ivan Laptev, Hossein Azizpour [KTH].

Deformable part-based models achieve state-of-the-art performance for object detection, but rely on heuristic initialization during training due to the optimization of non-convex cost function. This work investigates limitations of such an initialization and extends earlier methods using additional supervision. We explore strong supervision in terms of annotated object parts and use it to (i) improve model initialization, (ii) optimize model structure, and (iii) handle partial occlusions. Our method is able to deal with sub-optimal and incomplete annotations of object parts and is shown to benefit from semi-supervised learning setups where part-level annotation is provided for a fraction of positive examples only. Experimental results are reported for the detection of six animal classes in PASCAL VOC 2007 and 2010 datasets. We demonstrate significant improvements in detection performance compared to the LSVM and the Poselet object detectors.

This work has been published in [9].

6.2.3. Multi-Class Cosegmentation

Participants: Armand Joulin, Jean Ponce, Francis Bach [Inria SIERRA].

Bottom-up, fully unsupervised segmentation remains a daunting challenge for computer vision. In the cosegmentation context, on the other hand, the availability of multiple images assumed to contain instances of the same object classes provides a weak form of supervision that can be exploited by discriminative approaches. Unfortunately, most existing algorithms are limited to a very small number of images and/or object classes (typically two of each). This work proposes a novel energy-minimization approach to cosegmentation that can handle multiple classes and a significantly larger number of images. The proposed cost function combines spectral- and discriminative-clustering terms, and it admits a probabilistic interpretation. It is optimized using an efficient EM method, initialized using a convex quadratic approximation of the energy. Comparative experiments show that the proposed approach matches or improves the state of the art on several standard datasets.

This work has been published in [13].

6.2.4. A Convex Relaxation for Weakly Supervised Classifiers

Participants: Armand Joulin, Francis Bach [Inria SIERRA].

This work introduces a general multi-class approach to weakly supervised classification. Inferring the labels and learning the parameters of the model is usually done jointly through a block-coordinate descent algorithm such as expectation-maximization (EM), which may lead to local minima. To avoid this problem, we propose a cost function based on a convex relaxation of the soft-max loss. We then propose an algorithm specifically designed to efficiently solve the corresponding semidefinite program (SDP). Empirically, our method compares favorably to standard ones on different datasets for multiple instance learning and semi-supervised learning, as well as on clustering tasks.

This work has been published in [12].

6.2.5. Top-Down and Bottom-Up Cues for Scene Text Recognition

Participants: Karteek Alahari, Anand Mishra [IIT India], C.V. Jawahar [IIT India].

Scene text recognition has gained significant attention from the computer vision community in recent years. Recognizing such text is a challenging problem, even more so than the recognition of scanned documents. In this work, we focus on the problem of recognizing text extracted from street images. We present a framework that exploits both bottom-up and top-down cues. The bottom-up cues are derived from individual character detections from the image. We build a Conditional Random Field model on these detections to jointly model the strength of the detections and the interactions between them. We impose top-down cues obtained from a lexicon-based prior, i.e. language statistics, on the model. The optimal word represented by the text image is obtained by minimizing the energy function corresponding to the random field model.

We show significant improvements in accuracies on two challenging public datasets, namely Street View Text (over 15%) and ICDAR 2003 (nearly 10%).

This work has been published in [15].

6.2.6. Scene Text Recognition using Higher Order Language Priors

Participants: Karteek Alahari, Anand Mishra [IIT India], C.V. Jawahar [IIT India].

The problem of recognizing text in images taken in the wild has gained significant attention from the computer vision community in recent years. Contrary to recognition of printed documents, recognizing scene text is a challenging problem. We focus on the problem of recognizing text extracted from natural scene images and the web. Significant attempts have been made to address this problem in the recent past. However, many of these works benefit from the availability of strong context, which naturally limits their applicability. In this work we present a framework that uses a higher order prior computed from an English dictionary to recognize a word, which may or may not be a part of the dictionary. We show experimental results on publicly available datasets. Furthermore, we introduce a large challenging word dataset with five thousand words to evaluate various steps of our method exhaustively.

The main contributions of this work are: (1) We present a framework, which incorporates higher order statistical language models to recognize words in an unconstrained manner (i.e. we overcome the need for restricted word lists, and instead use an English dictionary to compute the priors). (2) We achieve significant improvement (more than 20%) in word recognition accuracies without using a restricted word list. (3) We introduce a large word recognition dataset (at least 5 times larger than other public datasets) with character level annotation and benchmark it.

This work has been published in [14].

6.3. Image restoration, manipulation and enhancement

6.3.1. Non-Uniform Deblurring for Shaken Images

Participants: Josef Sivic, Andrew Zisserman, Jean Ponce, Oliver Whyte [Microsoft Redmond].

Photographs taken in low-light conditions are often blurry as a result of camera shake, i.e. a motion of the camera while its shutter is open. Most existing deblurring methods model the observed blurry image as the convolution of a sharp image with a uniform blur kernel. However, we show that blur from camera shake is in general mostly due to the 3D rotation of the camera, resulting in a blur that can be significantly non-uniform across the image. We propose a new parametrized geometric model of the blurring process in terms of the rotational motion of the camera during exposure. This model is able to capture non-uniform blur in an image due to camera shake using a single global descriptor, and can be substituted into existing deblurring algorithms with only small modifications. To demonstrate its effectiveness, we apply this model to two deblurring problems; first, the case where a single blurry image is available, for which we examine both an approximate marginalization approach and a maximum a posteriori approach, and second, the case where a sharp but noisy image of the scene is available in addition to the blurry image. We show that our approach makes it possible to model and remove a wider class of blurs than previous approaches, including uniform blur as a special case, and demonstrate its effectiveness with experiments on synthetic and real images.

This work has been published in [8]. An image deblurring demo, described in section 5.8 , has been made available online.

6.3.2. Learning to Estimate and Remove Non-uniform Image Blur

Participants: Florent Couzinie-Devy, Jian Sun, Karteek Alahari, Jean Ponce.

This work addresses the problem of restoring images subjected to unknown and spatially varying blur caused by defocus or linear (say, horizontal) motion. The estimation of the global (non-uniform) image blur is cast as a multi-label energy minimization problem. The energy is the sum of unary terms corresponding to learned local blur estimators, and binary ones corresponding to blur smoothness. Its global minimum is found using Ishikawa's method by exploiting the natural order of discretized blur values for linear motions and defocus. Once the blur has been estimated, the image is restored using a robust (non-uniform) deblurring algorithm based on sparse regularization with global image statistics. The proposed algorithm outputs both a segmentation of the image into uniform-blur layers and an estimate of the corresponding sharp image. We present qualitative results on real images, and use synthetic data to quantitatively compare our approach to the publicly available implementation of Chakrabarti et al. 2010.

This work has been submitted to CVPR 2013.

6.4. Human activity capture and classification

6.4.1. Scene Semantics from Long-Term Observation of People

Participants: Vincent Delaitre, Ivan Laptev, Josef Sivic, David Fouhey [CMU], Abhinav Gupta [CMU], Alexei Efros [CMU].

Our everyday objects support various tasks and can be used by people for different purposes. While object classification is a widely studied topic in computer vision, recognition of object function, i.e., what people can do with an object and how they do it, is rarely addressed. In this work we construct a functional object description with the aim to recognize objects by the way people interact with them. We describe scene objects (sofas, tables, chairs) by associated human poses and object appearance. Our model is learned discriminatively from automatically estimated body poses in many realistic scenes. In particular, we make use of time-lapse videos from YouTube providing a rich source of common human-object interactions and minimizing the effort of manual object annotation. We show how the models learned from human observations significantly improve object recognition and enable prediction of characteristic human poses in new scenes. Results are shown on a dataset of more than 400,000 frames obtained from 146 time-lapse videos of challenging and realistic indoor scenes. Some of the estimated human poses and results of pixel-wise scene segmentation are shown in Figure 3 .

This work has been published in [10].

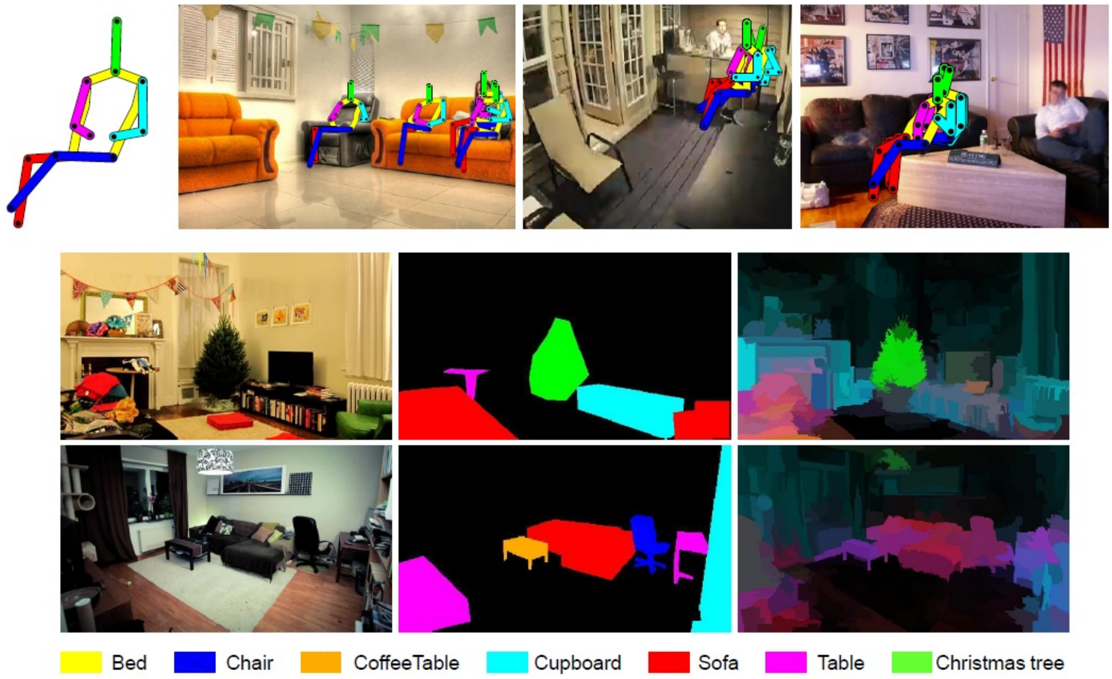


Figure 3. Top: Example of particular pose detections in three indoor scenes. Bottom: object segmentation illustrated by original images, ground truth segmentation, and automatic segmentation by our method shown in the left, middle and right columns respectively.

6.4.2. Analysis of Crowded Scenes in Video

Participants: Ivan Laptev, Josef Sivic, Mikel Rodriguez [MITRE].

In this work we first review the recent studies that have begun to address the various challenges associated with the analysis of crowded scenes. Next, we describe our two recent contributions to crowd analysis in video. First, we present a crowd analysis algorithm powered by prior probability distributions over behaviors that are learned on a large database of crowd videos gathered from the Internet. The proposed algorithm performs like state-of-the-art methods for tracking people having common crowd behaviors and outperforms the methods when the tracked individuals behave in an unusual way. Second, we address the problem of detecting and tracking a person in crowded video scenes. We formulate person detection as the optimization of a joint energy function combining crowd density estimation and the localization of individual people. The proposed methods are validated on a challenging video dataset of crowded scenes. Finally, the chapter concludes by describing ongoing and future research directions in crowd analysis.

This work is to appear in [17].

6.4.3. Actlets: A Novel Local Representation for Human Action Recognition in Video

Participants: Muhammad Muneeb Ullah, Ivan Laptev.

This work addresses the problem of human action recognition in realistic videos. We follow the recently successful local approaches and represent videos by means of local motion descriptors. To overcome the huge variability of human actions in motion and appearance, we propose a supervised approach to learn local motion descriptors – *actlets* – from a large pool of annotated video data. The main motivation behind our method is to construct action-characteristic representations of body joints undergoing specific motion patterns while learning invariance with respect to changes in camera views, lighting, human clothing, and other factors. We avoid the prohibitive cost of manual supervision and show how to learn actlets automatically from synthetic videos of avatars driven by the motion-capture data. We evaluate our method and show its significant improvement as well as its complementarity to existing techniques on the challenging UCF-sports and YouTube-actions datasets.

This work has been published in [16].

6.4.4. Layered Segmentation of People in Stereoscopic Movies

Participants: Karteek Alahari, Guillaume Seguin, Josef Sivic, Ivan Laptev.

In this work we seek to obtain a layered pixel-wise segmentation of multiple people in a stereoscopic video. This involves challenges such as dealing with unconstrained stereoscopic video, non-stationary cameras, complex indoor and outdoor dynamic scenes. The contributions of our work are three-fold: First, we develop a layered segmentation model incorporating person detections and pose estimates, as well as colour, motion, and stereo disparity cues. The model also explicitly represents depth ordering and occlusions of people. Second, we introduce a stereoscopic dataset with frames extracted from feature length movies “StreetDance 3D" and “Pina". In addition to realistic stereo image data, it contains nearly 700 annotated poses, 1200 annotated detections, and 400 pixel-wise segmentations of people. Third, we evaluate the benefits of stereo signal for person detection, pose estimation and segmentation in the new dataset. We demonstrate results on challenging realistic indoor and outdoor scenes depicting multiple people with frequent occlusions. Example result is shown in Figure 4 .

This work has been submitted to CVPR 2013.

6.4.5. Highly-Efficient Video Features for Action Recognition and Counting

Participants: Vadim Kantorov, Ivan Laptev.

Figure 4. A sample frame extracted from the stereoscopic movie "StreetDance": From left to right – left image from the stereo pair, disparity map computed from the stereo pair, layered segmentation of the image into 7 people. The front to back ordering is shown as a colour map, where "blue" denotes front and "red" denotes back. The cost function associated with our model is initialized using person detections, and incorporates disparity, pose, colour and motion cues. Note that the result shows accurate segmentation boundaries and also a reliable layer ordering of people.

Local video features provide state-of-the-art performance for action recognition. While the accuracy of action recognition has been steadily improved over the recent years, the low speed of feature extraction remains to be a major bottleneck preventing current methods from addressing large-scale applications. In this work we demonstrate that local video features can be computed very efficiently by exploiting motion information readily-available from standard video compression schemes. We show experimentally that the use of sparse motion vectors provided by the video compression improves the speed of existing optical-flow based methods by two orders of magnitude while resulting in limited drops of recognition performance. Building on this representation, we next address the problem of event counting in video and present a method providing accurate counts of human actions and enabling to process 100 years of video on a modest computer cluster.

This work has been submitted to CVPR 2013.

WIMMICS Team

6. New Results

6.1. Linked Data Access

Participants: Serena Villata, Luca Costabello, Fabien Gandon.

We designed and developed a context-aware access control framework for the Web of Data called Shi3ld [7]. The framework protects access to SPARQL endpoints, and it adopts Semantic Web languages only, as in the philosophy of the Web of Data. The innovative feature of the proposed framework consists in evaluating the accessibility to the data considering the attributes of the users. These attributes are defined following three main dimensions: user, device, and environment. The evaluation of the model shows that access control comes with a cost but it guarantees the protection of the data published on the Web of Data. The results of this research activity have been published in international conferences in the area of Artificial Intelligence (ECAI, [35]) and the Web (WWW [61], HT [34]).

On the same line, we have proposed a framework for attaching the licenses to the data resulting from a query on the Web of Data. The rationale is that the licenses associated to the data returned by the query are selected, and using a number of rules their compatibility is assessed. If the licenses are evaluated compatible, then they are composed into a composite license which is released to the user together with the data. The results of this research have been published in the COLD international workshop [74] and in the ISWC international conference [75] (best poster award). These two research lines have been performed in the context of the DataLift ANR project.

The PhD thesis of Luca Costabello, directed by F. Gandon and I. Herman (CWI and Semantic Web Activity Lead at W3C) investigates Web of Data interaction from mobile environments. Two main research activities have been carried out in 2012: i) PRISSMA [8], an adaptive rendering engine for RDF and ii) Shi3ld[9], a context-aware access control framework for Linked Data.

The goal of PRISSMA is delivering an *adaptive* rendering engine for Linked Data resources. PRISSMA tweaks RDF visualization to the context in which the resource consumption is performed. Work in 2012 has been mainly focused on determining the algorithm that selects the best RDF visualization according to the real, sensed context. The uncertain and incomplete nature of context data, led to investigate strategies that model the task as an inexact RDF instance matching operation.

The second main research line carried out in 2012 led to the creation of Shi3ld, an access control framework for Linked Data SPARQL endpoints. Shi3ld authorization procedure and Access Policies, defined using Semantic Web languages only, have been enriched with the notion of mobile context, thus enabling context-based access control (e.g. geo-temporal authorization policies).

A collateral research line has been carried out as the follow-up of the 9th edition of the Summer School on Ontology Engineering and the Semantic Web [10]. We contributed to an exploration work on the problem of spamming in Linked Data, providing a classification of potential spamming techniques and populating and publishing a dataset containing spammed triples[11]. The dataset is useful to train anti-spamming mechanisms.

6.2. ISICIL Platform

Participants: Nicolas Delaforge, Michel Buffa, Fabien Gandon, Alain Giboin.

[7] http://wimmics.inria.fr/projects/shi3ld/
[8] http://wimmics.inria.fr/projects/prissma
[9] http://wimmics.inria.fr/projects/shi3ld
[10] http://sssw.org/2012/
[11] http://www-sop.inria.fr/members/Luca.Costabello/spam/

ISICIL is an ANR project studying social networks and Semantic Web communities to support corporate intelligence.

First, ISICIL proposes a multidisciplinary design of a new form of corporate intelligence. The challenge of this project is to reconcile the new viral Web applications with formal representations of business processes and to integrate them into practical intelligence communities of actors working in the company. We designed, studied and experimented with new tools to support collaborative tasks in business intelligence by leveraging Web 2.0 interfaces (blog, wiki, social bookmarking) for interactions and Semantic Web technologies for interoperability and information processing. ISICIL also allowed to explore new scientific developments of the notion of epistemic cooperation (human interaction oriented toward the development and transmission of knowledge) and to identify usable technological solutions. An ergonomic approach, combining impregnation of ground truth data and freer technological inspirations from bibliographic and webographic sources, was proposed.

Secondly, ISICIL uses typed graphs as models underlying epistemic communities. The entire model relies on a unifying model based on RDF graphs to represent resources and community stakeholders. These models are integrated with bookmarking tools or "Web scraping" the results of which are tagged. The tags used are collected to form folksonomies and a semi-automatic thesaurus structure in these folksonomies. User feedback on this structure is captured when they use the search engine which offers tags related to their keywords and the user can accept, reject or adjust these suggestions. User profiles and links between them, considered as a network, are processed by a series of operators to propose a semantic analysis of social network, for example the computation of indicators of centrality parameterized by an ontology. Merged graphs of structured folksonomies and of social networks finally allow the detection and labeling of epistemic communities. Meanwhile we study how the analysis of user interactions can determine the trust and how to represent and control access to data and their semantics in a social network.

As part of the paradigm of Social Epistemology, ISICIL combined in an ergonomic approach, impregnation of ground truth data, a bibliometric study and technological inspirations to offer patterns and inferences exploiting the Semantic Web social networks to assist corporate intelligence. An open-source platform is available under CeCILL-C licence and was tested at ADEME and Orange. ISICIL is a proof of concept of the compatibility of the Semantic Web formalisms, practices and models of Web 2.0 and the philosophical framework of social epistemology.

The project has resulted in three PhD theses (Florie Bugeaud 2011, Guillaume Erétéo 2011 [98], Freddy Limpens 2010 [99]) and publications at the following conferences: ISWC 2009 IEEE / WIC / ACM Web Intelligence 2011, Web Science 2010, WWW 2011, ASWC 2009, COOP 2010, PRO-VE 2009, VSST 2010, EGC 2010, IC 2009 & 2011, Psycho Ergo Days 2010.

The ISICIL project participated to Colloque ANR STIC, January 4-6 2012.

6.3. Natural Language Processing

Participants: Elena Cabrio, Julien Cojan, Fabien Gandon, Maxime Lefrançois, Serena Villata.

We have proposed a combined framework of natural language processing and argumentation theory to support the users in their interactions within online debate platforms. The framework combines a natural language processing module which exploits the textual entailment approach and detects the arguments in natural language debates and the relationships among them, and an argumentation module which represents the debates as graphs and detects the accepted arguments. The results of this research have been published in one of the major conferences in the field of Artificial Intelligence (ECAI [30]), and in the areas of natural language processing (ACL [28]) and argumentation theory (COMMA [29]).

To enhance users interactions with the Web of Data, query interfaces providing an extensible mapping between natural language expressions, and concepts and relations in structured knowledge bases are becoming particularly relevant. As a result of the first year of her postdoc, Elena Cabrio (together with Julien Cojan) designed QAKiS (Question Answering wiKiframework-based System), that allows end users to submit a query in English to an RDF triple store and obtain the answer in the same language, hiding the complexity of the non intuitive formal query languages involved in the resolution process. At the same time, the expressiveness of these standards is exploited to scale to the huge amounts of available semantic data.

In its current implementation, QAKiS addresses the task of Question Answering (QA) over structured knowledge bases (e.g. DBpedia) where relevant information is expressed also in unstructured form (e.g. Wikipedia pages). Its major novelty is to implement a relation-based match for question interpretation, to convert the user question into a query language (e.g. SPARQL). A demo of the system is available online [12]. The results of this research have been published as a demo paper in the main conference of Semantic Web, ISWC [57], and in the special issue of the journal Intelligenza Artificiale [14]. QAKiS has been evaluated with respect to state-of-the-art systems taking part into the QALD-2 (Question Answering over Linked Data) challenge at ESWC, obtaining satisfactory results [59].

In the PhD Thesis of Maxime Lefrançois, we are interested in bridging the world of natural language and the world of the Semantic Web in particular to support multilingual access to the Web of Data, and management of interlingual knowledge bases. In 2011 we introduced the ULiS project that aimed at designing a pivot-based NLP technique called Universal Linguistic System, 100% using the Semantic Web formalisms, and being compliant with the Meaning-Text theory [102].

We showed that neither Description Logics nor Conceptual Graphs suit our needs, so this Ph.D. now focuses on the formalization of the Unit Graphs mathematical framework that is conceived to fill the gap between the highly linguistically precise Explanatory Combinatorial Dictionaries of the Meaning-Text Theory and the Dependency Grammars, and the highly mathematically grounded model of the Conceptual Graphs.

Maxime finally joined the Multilingual-Web-LT W3C Working Group, and left it when the charter got revised.

6.4. Requirement Engineering

Participants: Zeina Azmeh, Isabelle Mirbel, Serena Villata.

Requirements engineering is an essential process of software engineering, during which the complete behavior of a software system can be defined. The success of this process plays a crucial role in the success of the whole software project. A key issue of requirements engineering is stakeholders participation, which is facilitated through the emergence of online collaborative working tools. These tools create new opportunities of practice regarding requirements elicitation. Nevertheless, they result in an information overload lacking structure and semantics. Consequently, requirements analysis and selection become more challenging.

Our current proposition is embodied in an approach based on Semantic Web languages as well as concept lattices to identify relevant communities of stakeholders depending on their past participation. These communities can be used to enable efficient decision-making and handling of requirements.

We exploited the idea of applying argumentation theory to deal with requirements engineering. In particular, the proposed framework detects consistent sets of goal-based requirements and maintains their consistency over time based on argumentation theory which allows to detect the conflicts among elements. More specifically, the framework relies on meta-argumentation, which instantiates abstract argumentation frameworks, where requirements are represented as arguments and the standard Dung-like argumentation framework is extended with additional relations between goal-based requirements. The results of this research have been published to the CLIMA international workshop [37].

6.5. Regulation Engineering

Participants: Khalil Bouzidi, Catherine Faron-Zucker, Olivier Corby.

Regulations in the Building Industry are becoming increasingly complex and involve more than one technical area, covering products, components and project implementations. They also play an important role in ensuring the quality of a building, and to minimize its environmental impact.

[12] http://dbpedia.inria.fr/qakis/

In a collaboration between CSTB and the I3S laboratory, we are carrying on research on the acquisition of knowledge from the technical and regulatory information contained in the REEF referential [13] and the automated processing of this knowledge with the final goal of assisting professionals in the use of these texts and the creation of new texts.

We are applying this work in CSTB to help industrials in the writing of Technical Assessments. The problem is how to specify these assessments and standardize their structure using models and adaptive semantic services.

A Technical Assessment (in French: Avis Technique ou ATec) is a document containing technical information on the usability of a product, material, component or element of construction, which has an innovative character. We chose this Technical Assessment as a case study because CSTB has the mastership and a wide experience in these kinds of technical documents.

In 2012, we were particularly interested in the modeling of the regulatory constraints derived from the Technical Guides used to validate the Technical Assessment. These Guides are regulatory complements offered by CSTB to the various industrials to enable easier reading of technical regulations. They collect execution details with a wide range of possible situations of implementations.

Our work aims to formalize the Technical Guides in a machine-processable model to assist the creation of Technical Assessments by automating their validation.

Our first contribution is the use of standard SBVR (Semantics of Business Vocabulary and Business Rules) and SPARQL to reformulate the regulatory requirements of guides on the one hand in a controlled language and on the other hand in a formal language

Second, our model incorporates expert knowledge on the verification process of Technical Documents. We have organized the SPARQL queries representing regulatory constraints into several processes. Each component involved in the Technical Document corresponds to an elementary process of compliance checking. An elementary process contains a set of SPARQL queries to check the compliance of an elementary component. A full complex process for checking a Technical Document is defined recursively and automatically built as a set of elementary processes relative to the components which have their semantic definition in the OntoDT ontology that we have designed.

Finally, we represent in RDF the association between the SBVR rules and SPARQL queries representing the same regulatory constraints. We use annotations to produce a compliance report in natural language to assist users in the writing of Technical Assessments.

As a result, we have designed a Semantic Web application to support and guide the process of writing Technical Assessment. The current version has allowed us to validate our approach. Also, we have developed a base of SBVR rules to describe business requirements of guides. This rule base is implemented in SPARQL.

6.6. Graph-based Knowledge Representation

Participants: Olivier Corby, Catherine Faron-Zucker, Fabien Gandon, Isabelle Mirbel, Adrien Basse, Oumy Seye.

We have designed a method to build pretty-printers for RDF Abstract Syntax Trees and Graphs, built on top of SPARQL Query Language. Pretty-print rules are written as SPARQL select-where queries. The *where* clause matches the target subtree to be printed and the *select* clause returns the pretty-printed statement using an external *kg:pprint* function. This function recursively calls the pretty printer, looking for appropriate pretty-print queries for the target subtrees.

We have designed a syntactic extension to SPARQL in order to ease writing pretty-printing rules. Below, an example of template for a SPIN like AST is shown:

```
template {
  "select " ?s
```

[13] http://reef.cstb.fr

```
  "\n"
  "where "  ?w
}
where {
  ?in ast:select ?s ;
      ast:where  ?w
}
```

We were able to write a pretty-printer for a SPIN like complete SPARQL 1.0 AST with 19 templates.

We have designed an extension to our KGRAM SPARQL interpreter that enables to consider an RDF Graph directly as a Query Graph. This enables to compute RDF subgraph matching.

We have completed SPARQL 1.1 implementation with the final version of Property Path (PP) and federated queries (service & bindings). In order to be able to query a SPARQL 1.0 endpoint with PP, we have written a compiler that translates PP into basic graph patterns.

The work on KGRAM is published in [33].

Alban Gaignard from the I3S Modalis team has designed a distributed version of KGRAM to query remote triple stores, in the context of Semantic Federation of Distributed Neurodata. This work is published in [63], [64].

We have implemented a prototype of C-SET Commutative Replicated Data Type for RDF in Corese with Pascal Molli and Luis Ibáñez from U. Nantes [69].

6.6.1. Extracting Graph Patterns to Characterize RDF Data Sources

This work takes place in the PhD Thesis of Adrien Basse.

Many Semantic Web applications address the issue of integrating data from distributed RDF triple stores. There are several solutions for distributed query processing such as SPARQL 1.1 Federation, which defines extensions to the SPARQL Query Language to support distributed query execution. Such extensions make it possible to formulate a query that delegates parts of the query to a series of services, but one issue remains: how to automate the selection of RDF triple stores containing relevant data to answer a query. This is especially true in the context of the Linking Open Data project where numerous and very heterogeneous datasets are interlinked, allowing for interesting queries across several sources. To decompose and send queries targeting only relevant stores, we need a means to describe each RDF triple store, i.e. an index structure which provides a complete and compact description of the content of the RDF triple store.

To know the content of a RDF triple store, we proposed to use graph patterns as basic structures for index items. In this thesis we present an approach to extract these graph patterns from RDF triple store. For this purpose, we extend Depth-First Search coding (DFS) [104] to RDF labeled and oriented multigraphs and we provide a join operator between two DFS codes so as to sequentially build the different levels of the index structure.

Insertion or deletion of annotations in the triple store may cause changes to the index structure. To handle updates in triple store, we proposed a procedure to identify exactly the changes in the first level of the index structure and propagate them to the following levels. The DFSR (Depth First Search for RDF) coding makes it possible for us to efficiently manipulate graph patterns, but is difficult to read (succession of integer numbers). To facilitate the reading of our index structure, we propose a visualization user-interface and algorithms to turn a DFS code into a more legible format like RDF. Our algorithm relies on Corese/KGRAM [95]. We have tested our algorithm on many datasets. During the building of index structures we keep a set of data in order to help us to better understand the progress of our algorithm and improve it.

6.6.2. Rules for the Web of Data

This work takes place in the PhD Thesis of Oumy Seye.

We have characterized the subset of SPARQL that can be expressed in RIF and, conversely, we have searched for the maximal RIF dialect that can be expressed as SPARQL construct-where queries. This work is published in [71], [72] were we present the implementation of a RIF dialect with a SPARQL Rule Engine in Corese/KGRAM.

We have designed online services for RIF-BLD parsers for presentation syntax and XML syntax [14]. We have also done an online service for RIF-BLD translation into SPARQL and RDF [15].

6.7. Business Intelligence

Participants: Corentin Follenfant, Olivier Corby, Fabien Gandon.

This PhD Thesis is done with a CIFRE industrial grant from SAP Research.

Industrial Business Intelligence proposes tools and methods to perform data analysis over heterogeneous enterprise sources. They allow one to harvest, federate, cleanse, annotate, query, organize and visualize data in order to support decision making with human-readable documents such as reports, dashboards, mobile visualizations. Authoring these dynamic documents requires proficiency in technical domains like relational modeling and SQL for one to produce relevant content: end users therefore praise example-driven and information retrieval (IR) systems that help them reusing existing content. Such systems need common structured metadata to enable comparison, search, matching and recommendation of (parts of) documents.

As target data sources are mainly tabular or relational, queries executed to feed the dynamic documents are SQL or derivatives. In [62] we proposed to model these queries as RDF named graphs, and use the graphs as documents annotations. Queries are represented through their abstract syntax trees (AST) represented with RDF graphs. The SQL-specific modeling contribution can therefore be applied to any generic query language. We identified two desirable features for IR systems that deal with queries repositories: search and rewriting, the latter allowing further annotation as well as reconciliation of source relational entities against LOD (Linked Open Data) repositories. On this basis we evaluated SPARQL 1.1 to perform SQL query analysis, i.e. pattern-matching search or rewriting, using in particular property paths. Resulting SPARQL queries are intuitive and concise.

Next steps include a quantitative evaluation by extracting RDF representations from a repository of SQL-fed documents, the production of a library of SPARQL queries that perform generic IR operations against RDF-modelled SQL queries, a formalization of the modeling and operations to compare them with generic tree manipulation methods. In further work we plan to investigate rewriting queries from different languages modelled with language-specific abstract syntax trees to generic abstract syntax trees and experiment cross-language query comparison with SPARQL.

6.8. Fuzzy Knowledge Representation

Participant: Andrea Tettamanzi.

Andrea Tettamanzi has joined the Wimmics research team in September 2012, after winning a position as a full professor at the University of Nice-Sophia Antipolis. He got a PhD in Computational Mathematics and Operations Research in 1995 from the University of Milan with a thesis on evolutionary algorithms; he became assistant professor at the University of Milan in 1998 and associate professor at the same university in 2002.

His research interests focus on combining different methods of computational intelligence, namely evolutionary algorithms, fuzzy logic, and neural networks, to solve real-world problems in the presence of imprecision, noisy data, and ill-defined optimization criteria, but also on the management of vagueness and uncertainty in knowledge representation, the automatic extraction of knowledge from data, possibility theory and its application to belief revision and goal generation in cognitive agents.

[14] http://wimmics-ws.inria.fr/rifparser
[15] http://wimmics-ws.inria.fr/riftosparql

After joining Wimmics, Andrea Tettamanzi has continued work on previous collaborations with other members of the team, namely Serena Villata [76], and has begun exploring with the rest of the team several research axes that could benefit from his contribution.

6.9. Emotion Modeling

Participants: Franck Berthelon, Imen Tayari, Nhan Le Thanh, Peter Sander.

In the PhD Thesis of Imen Tayari, an algebraic vector representation model of emotional states was designed. This multidimensional model provides a powerful mathematical tools for the analysis and the processing of emotions. It permits to integrate information from different modalities(speech, facial expressions, gestures) in order to allow more reliable estimation of emotional states. Indeed our proposal aims at efficient recognition of emotional states even when they appear to be superposed or masked. Experiments show the efficiency of the proposed method in detecting basic emotion by giving hight recognition rate. This work is published in [39], [41], [43], [40], [42].

In the PhD Thesis of Franck Berthelon, we are working in the domain of affective computing to create an emotion sensitive system. Interaction between human and computer can be improved via such a system because emotion is so important in everyday communication. Our research focuses on serious gaming, particularly on enabling a user and a virtual character to "share" an emotion. The two main problems that arise are:

- How to detect a user's emotions given that the stimulus comes from a virtual environment?
- How to give feedback based on the user's current emotion?

We propose to model emotions as a complex system where data are retrieved from physiological sensors such as heart rate, EMG or EEG. We need to map the multi-sensor data back into a dimensional model of emotion space. Finally, we aim to have an effect on the user's emotional state by varying the stimulus received from the virtual environment. This puts the user into different emotional situations determined by the task to accomplish, with an accompanying effect on their ability to carry out the task.

We developed an application for experimentation purposes; it aims to implement our model using an EEG headset (Emotiv EPOC). This application allows us to generate an emotional map from a slide show of emotion annotated picture. Based on the created map and the real time EEG data, this application can compute a user's instantaneous emotion.

In addition to the first development, we reuse experimental data from MIT to validate our model in a more controlled way. We take the same data, features, signal processing and feature reduction algorithm but instead of using the k-nearest neighbors (KNN) classification algorithm we use our model to identify and annotate discontinuities that represents emotion state changes in accordance with Klaus R. Scherer hypothesis [103].

We are continuing work on validating our model with quantitative results and applying those results on a more realistic system with the application we have developed.

6.10. DBpedia in French

Participants: Julien Cojan, Fabien Gandon.

The purpose of the project *DBpedia in French* is to extract data from Wikipedia in French and publish it under structured format. Wikipedia content is mainly meant to be read by human and is not suited for use in applications. DBpedia publishes the data extracted from Wikipedia articles in RDF W3C standard for the Semantic Web [16] thus readily available for many applications. For instance, DBpedia is used to generate indexes for cultural resources (e.g. HdA-lab project [17]), it can also be used for mobile applications thanks to the geographic data it contains, or to answer natural language questions, etc.

[16] http://www.w3c.org/RDF/
[17] http://hdalab.iri-research.org/hdalab/

The original version of DBpedia is focused on the English chapter of Wikipedia. Last versions also contain elements extracted from other chapters, but only when related to a page in English. Articles with no equivalent in English are skipped, leading to a significant number of pages being ignored and so a significant amount of data is lost. For instance, about 49 000 persons and 180 000 places described in the French chapter have no corresponding article in English and are then missing in the English DBpedia. Moreover, the description of the same topic can be different from one chapter to another, reflecting cultural diversity.

DBpedia in French publishes data extracted from the French Wikipedia in complement to the English DBpedia. Data are linked with the different chapters from the internationalization committee thus providing multilingual resources. In its release from October 2nd, DBpedia in French contains 130 million triples describing 1.3 million things, among them 260 000 places, 140 000 persons, 64 000 work pieces and 26 000 organizations.

This project is supported by the Semanticpedia collaboration platform [18] launched November 19th 2012 by Aurélie Filipetti, the French Ministry of Culture, Michel Cosnard, CEO of Inria, and Rémi Mathis, CEO of Wikimédia France. Inria currently hosts the project [19] and is the correspondent for the French chapter in DBpedia internationalization committee.

6.11. Co-Construction of Community Ontologies

Participants: Papa Fary Diallo, Isabelle Mirbel, Olivier Corby.

PhD Thesis on Co-Construction of Community Ontologies and Corpus in a Limited Technological Environment.

To refresh the memory of people and revive many stories that accompany the creation and daily life of different Africa territories, the establishment of an online sociocultural encyclopedia was conceived. It will be an online platform based on a Geographical Information System (GIS) enriched by a semantic layer allowing access to different information.

In the last decade, we have seen the rise of two visions of the Web: on one side the Social Web or Web 2.0, which places users at the heart of the Web, they are no longer spectator but become editor of the content of Web pages. On another hand, the Semantic Web proposes knowledge representations (ontologies) that allow machines (software agent) to better understand data on the Web. Both aims were often opposed, but there are a lot of work trying to combine these visions.

In our work, oriented in this direction, we will try to create a new point of view of Community concept within the Web. Community is a group of people who share a common set of values and interests. This shift of view allows us to address a specific community as an atomic entity and focus this time on the sharing of knowledge between communities. The second challenge is to combine Social Web and Semantic Web technologies. Using Semantic Web in our social network, we have a semantic layer that provides access to various information contained in the network. Furthermore, the Semantic Web opens up a semantic approach to social network analysis, which also allows extracting new knowledge.

In this thesis, we study the implementation of an online platform to build and share the collective memory of citizens and revive many stories by using a semantic layer. Semantic Social Network Analysis will allow us to present data in eye-catching way and in another view. This platform will be updated by the actors and the citizens of these territories, and share their history and heritage through their "social network".

6.12. Semantic Wiki

Participants: Pavel Arapov, Michel Buffa.

[18] http://semanticpedia.org
[19] http://fr.dbpedia.org

A Wiki is a Web site that lets users create and edit content collaboratively. A Wiki engine is a programming base to create Wiki sites. In this PhD Thesis, our approach to the creation of a Wiki engine is to use an application on the Wiki pages and semantic meta-data. Our vision for the Wiki pages is that a Wiki page is an application. We do not work with static data on the Internet neither in a Wiki, now it is a Web application that contains the source code of the application interface, as well as the data for display. Application is able to retrieve and update data based on Linked Data principles on each page load, updating their sources as needed and showing only relevant information. This work is published [51], [52].

6.13. Discovery Hub

Participants: Nicolas Marie, Fabien Gandon, Damien Legrand.

Nicolas Marie is PhD student in collaboration with Alcatel-Lucent Bell Labs (Cifre). He is the project leader of Discovery Hub: a discovery engine on the top of DBpedia using real-time spreading activation.

We continue the CRE and CIFRE PhD Thesis (2011-2013) initially on Social objects, object-centered sociality, and object-centered social networks to propose mobile context-based notification application in a semantic and pervasive Web. The work evolved toward exploratory search, discovery and recommendation. Web growth, both in size and diversity, and users' growing expectations increase the need for innovative search approaches and technologies. Exploratory search systems are built specifically to help user in cognitive consuming search tasks like learning or investigation. Some of these systems are built on the top of linked data and use its semantic richness to provide cognitively-optimized search experiences. This work addresses the question of real-time linked data processing for exploratory search purposes. This live aspect offers advantages in terms of query expressivity and data dynamicity-handling.

To achieve this goal we propose a real-time semantic spreading activation algorithm (RTSA) which process linked data on-the-fly. This live aspect offers advantages in data dynamicity handling and query expressivity. Approximation strategies, algorithm behavior study and user evaluation related to RTSA algorithm are currently performed. The work includes a study of its behavior on DBpedia and a validation of its relevance through a user evaluation. Finally we also implemented a real deployment introducing the Discovery Hub prototype. It is an exploratory search engine offering advanced querying, browsing and explanation strategies for discovery purposes.

This algorithm is deployed in the Discovery Hub prototype [20], a discovery engine interfaced with services. Discovery Hub aims to help users to make numerous discoveries starting from its interests. The application works on DBpedia data including local version like fr.dbpedia.org (hosted by Inria/Wimmics). It also makes extensive use of the Corese/KGRAM Semantic Web Factory. Application front-end was designed and developed by Damien Legrand during an internship.

The application was presented during the Semanticpedia day, organized by official French language delegation [21].

Nicolas Marie is also active in the Web Science community [36].

6.14. Semantic Newsfeed Aggregation

Participant: Christophe Desclaux.

Christophe was this year in his last year of engineering school at Polytech UNS in the KIS speciality. During his end of course project he worked on the RSS feed aggregation using Named Entities Recognition. He presented his research project to the student contest *Boost Your Code* organized by Inria. The aim of the contest is to offer to a junior engineer a one year full time contract to work on an innovating OpenSource project. Christophe won the contest and is now part of the Wimmics team since november 2012. The ZONE project [22] provides a specialized tool for monitoring domain. ZONE semantically increases news for a better classification for the user. Christophe will work in collaboration with the team on documents clustering, natural language processing and RDF datastores.

[20] http://semreco.inria.fr/
[21] http://www.dglflf.culture.gouv.fr/Actualites/Programme_Semanticpedia.pdf
[22] http://zone-project.org

6.15. Linked Justifications

Participants: Rakebul Hasan, Fabien Gandon.

Semantic Web applications use inferential capabilities and distributed data in their reasoning. Users often find it difficult to understand how these applications produce their results. Hence, users often do not trust the results these applications produce. Explanation-aware Semantic Web applications provide explanations of their reasoning. Explanations enable users to better understand reasoning of these application. Users can use this additional information about reasoning to make their trust decisions.

The emergence of Linked Data offers opportunities for large-scale reasoning over heterogeneous and distributed data. Explaining reasoning over Linked Data requires explaining how these distributed data were produced. Publishing also the explanation related metadata as Linked Data enables such explanations. Justifications are metadata about how a given piece datum is obtained. We introduce the concept of Linked Justifications and provide guidelines to publish justifications as Linked Data in [67]. We published the *Ratio4TA*[23] (inte**r**linked justific**atio**ns **for** **t**riple **a**ssertions) vocabulary to describe justifications. *Ratio4TA* extends W3C PROV Ontology[24] to promote interoperability.

In [89], [66], we analyze the existing explanation-aware Semantic Web systems. The existing systems inherit explanation features from explanation-aware expert systems. These explanations are targeted to expert users, such as knowledge engineers, with detailed information about all the execution steps of reasoners of these applications. Unlike the expert systems, users of Semantic Web applications have diverse background - from expert knowledge engineers who are interested in every details of the reasoning, to regular users who do not have any background in reasoning, logic, or ontologies. These non-expert users might feel overwhelmed with all the execution details of reasoners. To address this issue, we propose summarized and relevant explanations to users. Users can specify their explanation goals - types of information they are interested in. We take into consideration the explanation goals when we present explanations and summarize explanations. We use centrality and similarity matrices to summarize and provide relevant explanations.

6.16. Analyzing and Modeling Users, Communities and their Interactions in a Social Semantic Web Context

6.16.1. Analyzing and Modeling the Sharing and Articulation of Representations

Participants: Alain Giboin, Gessica Puri.

Comparing and Bridging Models of Representation Sharing Processes

Context: Follow-up to the RefCom joint research action of the GDR CNRS Psycho-Ergo, in collaboration with Pascal Salembier (UTT, France).

We continued our work on comparing and bridging models of representation sharing processes (see Edelweiss activity report 2011) in order to achieve mutual intelligibility between researchers working on such models. We extended at the European level the test and application of the grid we elaborated for collaboratively comparing and bridging the conceptualizations [86], [65].

Methods and tools for articulating developers, domain experts, users and usage analysts' representations

Reconciling informal and formal representations through the ECCO collaborative ontology editor

Context: Follow-up to the ANR project e-WOK HUB, in collaboration with Priscille Durville (a former Inria expert engineer, currently engineer at Xerox), Sandrine Grataloup and Olivier Morel (BRGM), Michel Perrin (ENSMP)

[23] http://ns.inria.fr/ratio4ta/
[24] http://www.w3.org/TR/prov-o/

In a new publication [81], we reported the method we designed and proposed to geologists and knowledge engineers to help them jointly define domain ontologies from textual documents. The method is instrumented through a collaborative ontology editor (ECCO) which integrates two tools for automatic analysis of natural language. ECCO allows articulating the informal representations formulated by geologists in natural language and the knowledge engineers' fully formalized ontology-based representations that can be processed automatically by a semantic search engine like Corese.

Frameworks and Toolsets

Context 1: Capitalizing methods and tools developed in the Wimmics (formerly Edelweiss) team

We started to integrate into a general framework the methods and tools developed in the Wimmics (formerly Edelweiss) team to help developers, domain experts, users and usage analysts reconcile their views in order to design user-adapted social semantic applications. A preliminary presentation of the framework was given in the interdisciplinary seminar mentioned in Section "Invited Talks".

Context 2: PhD thesis of Gessica Puri and Wimmics projects related to visualization and manipulation of links.

We are currently developing a "design thinking" toolset (including a framework) for helping developers think in terms of a user's point of view when they design and evaluate link visualization and manipulation applications such as graph visualization applications [90]. A first version of the toolset is being validated by developers of the team.

6.16.2. Scenario Modeling and User Modeling for system design and evaluation

Participants: Alain Giboin, Gessica Puri.

In the context of different design projects, we applied, adapted or renewed some of our scenario and user modeling methods.

Context 1: ISICIL project. In collaboration with Rubiela Silva (UNS & Ademe), Claire Prendleloup (Ademe), Mylène Leitzelman (Telecom ParisTech)

In order to evaluate the usefulness and usability of the ISICIL platform (see 6.2) for one of the communities of potential users of the platform (the French Environment and Energy Management Agency), we adapted in particular: (a) a technique for modeling collectives and their related scenarios to prepare testing situations as close as possible to real situations met by the potential users; and (b) an existing set of collective heuristics, i.e., heuristics formerly designed to evaluate groupware; see, e.g. [93]. We also proposed the technique of *Online collaborative sessions* supported by the ISICIL communication functionalities (chat and comments) as a writing-based version of the *think out loud* protocol. These techniques are partly described in [91], [92].

Context 2: PAL project, in collaboration with David Daney (Coprin), Rémi Barraquand (Prima), Nadine Mandran (Pôle d'Ingénierie Multidisciplinaire du Laboratoire d'Informatique de Grenoble)

In this work we applied and adapted techniques for articulating system-oriented scenarios (coming from robotics and computer vision specialists) and user-oriented scenarios. These scenarios were intended to guide the design of useful and usable services improving the autonomy and quality of life for elderly and fragile persons. An output of this work has been the organization of workshop where PAL researchers were invited to specify the scenarios motivating the development of their services.

Context 3: Projects *Socio-cultural encyclopedia of Senegalese communities* and *Global Warming Platform* – Papa Fary Diallo (Wimmics PhD student), Fatou Kamara and Moussa Lo (Université Gaston-Berger, Saint-Louis, Sénégal)

We started to apply techniques for modeling groups of users and their related scenarios to the design of social semantic applications aimed at (a) communities wanting to adapt these applications to their own culture, and (b) at communities including persons from different specialties (geographers, mathematicians) and laypersons which want to collaborate on global warming issues.

Context 4: Discovery Hub project, PhD thesis of Nicolas Marie, in collaboration with Florentin Rodio (Alcatel Lucent); also related to the PhD thesis of Gessica Puri.

In order to perform a user-oriented evaluation of the Discovery Hub recommender system (see 6.13 and [100]), we used techniques allowing to define more realistic scenarios of interaction with the recommender system from the user's point of view, and to analyze users' cognitive processes when they interact with the system (e.g., when they select recommended items or when they assess the quality of a recommendation with and without explanations).

6.16.3. Exploring eye-tracking techniques for evaluating information organization aspects in Web applications

Participants: Valériane Dusaucy, Alain Giboin.

Context: collaboration with Valériane Dusaucy (PhD student, University of Aix-Marseille & CIFRE Société Ausy) and Franck Ferront (ergonome, Société Ausy)

We designed an experiment to explore the potentiality of eye-tracking techniques for evaluating information organization aspects in Web applications from a user's point of view, and to compare it to other evaluation techniques such as heuristic evaluation. The experiment, which takes place in the Ubiquarium of the I3S Laboratory, is in progress.

ZENITH Project-Team

6. New Results

6.1. Data and Metadata Management

6.1.1. Uncertain Data Management

Participants: Reza Akbarinia, Patrick Valduriez, Guillaume Verger.

Data uncertainty in scientific applications can be due to many different reasons: incomplete knowledge of the underlying system, inexact model parameters, inaccurate representation of initial boundary conditions, inaccuracy in equipments, error in data entry, etc.

One of the areas, in which uncertainty management is important, is the integration of heterogeneous data sources, in the sense where usually there may be an uncertainty in the possible mappings between the attributes of the sources. Usually the human interaction is demanded to help the system in choosing the correct mappings. In [30], we propose a pay-as-you-go data integration solution that aims at preforming the data integration in a fully automated way. Our solution takes advantage of attribute correlations by using functional dependencies, and captures uncertainty in mediated schemas using a probabilistic data model. It allows integrating a given set of data sources, as well as incrementally integrating additional sources, without needing to restart the process from scratch. We implemented our solution, and compared it with a baseline approach. The performance evaluation results show significant performance gains of our solution in terms of recall and precision compared to the baseline approaches.

Another problem that arises in many applications such as data integration systems is that of Entity Resolution (ER). ER is the process of identifying tuples that represent the same real-world entity. It has been well studied in the literature for certain data, but it has not been deeply investigated for uncertain data. Existing proposals for the ER problem are not applicable to the above examples since they ignore probability values completely and return the most similar tuples as the solution. Furthermore, the semantics of the solution for the ERUD problem has not been clearly defined in the literature. In [31], we address the ERUD problem. We adopt the well-known possible worlds semantics for defining the semantics for the ERUD problem, and propose a PTIME algorithm for a large class of similarity functions, i.e. context-free. For the rest of similarity functions, i.e. context-sensitive, we use Monte-Carlo randomization for approximating the answer. We propose a parallel version of our Monte-Carlo algorithm using the MapReduce framework. To the best of our knowledge, this is the first study of the ERUD problem that adopts the possible world semantics and the first efficient algorithm for implementing it.

Another important problem in uncertain data management is the efficient processing of probabilistic queries. We have continued the development of our probabilistic database prototype, called ProbDB (Probabilistic Database) that deals with large-scale probabilistic data sharing. ProbDB divides each probabilistic query into two parts: probabilistic and deterministic (i.e. non probabilistic). The deterministic part is executed by the underlying RDBMS, and the rest of work is done by our probabilistic query processing algorithms that are executed over the data returned by the RDBMS.

6.1.2. Metadata Integration

Participants: Zohra Bellahsène, Emmanuel Castanier, Duy Hoa Ngo, Patrick Valduriez.

Our work on metadata integration encompassed ontology matching and open data source integration.

The major focus of our work in 2012 was to deal with large scale ontology matching and scalability. To improve the matching quality of YAM++, we designed a new IR-based measure to deal with terminological heterogeneity in real world ontologies. To deal with large ontology matching, we designed a method based on indexing concepts from their labels and comments. Our approach aims at reducing the search space when comparing the concepts of the input ontologies. For this purpose, we designed three filters: Description Filter, Context Filter and Label Filter. These methods make use of the Lucene search engine for indexing and searching the context of entities in the input ontologies. Another contribution lies on the Fast Semantic Filtering method, which refines the discovered mappings in the ontology matching task. The aim of the Semantic Filter is to detect and reject inconsistent mappings by exploring semantic information of entities in the input ontologies [45]. The originality of our method is to use a new structural indexing technique and a heuristic to generate relative disjointness axioms. At the 2012 competition of the Ontology Alignment Evaluation Initiative (http://oaei.ontologymatching.org), YAM++ was one of the best matchers, with very good results in all tracks. It obtained the first postision in the Large BioMed Track [55].

Integrating open data sources can yield high value information but raises major problems in terms of metadata extraction, data source integration and visualization of integrated data. In [34], [33], we describe WebSmatch, a flexible environment for Web data integration, based on a real, end-to-end data integration scenario over public data from Data Publica. WebSmatch supports the full process of importing, refining and integrating data sources and uses third party tools for high quality visualization. We use a typical scenario of public data integration which involves problems not solved by currents tools: poorly structured input data sources (XLS files) and rich visualization of integrated data.

6.1.3. High-dimensional data management

Participants: Mohamed Riadh Trad, Alexis Joly, Saloua Litayem.

High dimensional data hashing is essential for scaling up and distributing data analysis applications involving feature-rich objects, such as text documents, images or multi-modal entities (scientific observations, events, etc.). In this first research track, we first investigated the use of high dimensional hashing methods for efficiently approximating K-NN Graphs [47], particularly in distributed environments. We highlighted the importance of balancing issues on the performance of such approaches and show why the baseline approach using Locality Sensitive Hashing does not perform well. Our new KNN-join method is based on RMMH, a hash function family based on randomly trained classifiers that we introduced in 2011. We show that the resulting hash tables are much more balanced and that the number of resulting collisions can be greatly reduced without degrading quality. We further improve the load balancing of our distributed approach by designing a parallelized local join algorithm, implemented within the MapReduce framework. In other work [43], we address the problem of speeding-up the prediction phase of linear Support Vector Machines via Locality Sensitive Hashing. Whereas the mainstream work in the field is focused on training classifiers on huge amount of data, less efforts are spent on the counterpart scalability issue: how to apply big trained models efficiently on huge non annotated collections ? In this work, we propose building efficient hash-based classifiers that are applied in a first stage in order to approximate the exact results and alter the hypothesis space. Experiments performed with millions of one-against-one classifiers show that the proposed hash-based classifier can be more than two orders of magnitude faster than the exact classifier with minor losses in quality.

6.2. Data and Process Sharing

6.2.1. Hybrid P2P/cloud Architecture

Participants: Esther Pacitti, Patrick Valduriez.

Zenith adopts a hybrid P2P/cloud architecture. P2P naturally supports the collaborative nature of scientific applications, with autonomy and decentralized control. Peers can be the participants or organizations involved in collaboration and may share data and applications while keeping full control over some of their data (a major requirement for our application partners). But for very-large scale data analysis or very large workflow activities, cloud computing is appropriate as it can provide virtually infinite computing, storage and networking resources. Such hybrid architecture also enables the clean integration of the users' own computational resources with different clouds.

In [24], we define Zenith's architecture with P2P data services and cloud data services. We model an online scientific community as a set of peers and relationships between them. The peers have their own data sources. The relationships are between any two or more peers and indicate how the peers and their data sources are related, e.g. friendship, same semantic domain, similar schema. The P2P data services include basic services (metadata and uncertain data management): recommendation, data analysis and workflow management through the Shared-data Overlay Network (SON) middleware. The cloud P2P services include data mining, content-based information retrieval and workflow execution. These services can be accessed through web services, and each peer can use the services of multiple clouds.

6.2.2. Social-based P2P Data Sharing

Participants: Reza Akbarinia, Emmanuel Castanier, Esther Pacitti, Didier Parigot, Patrick Valduriez, Guillaume Verger.

As a validation of the ANR DataRing project, we have developed P2PShare, a P2P system for large-scale probabilistic data sharing in scientific communities. P2PShare leverages content-based and expert-based recommendation. It is designed to manage probabilistic and deterministic data in P2P environments. It provides a flexible environment for integration of heterogeneous sources, and takes into account the social based aspects to discover high quality results for queries by privileging the data of friends (or friends of friends), who are expert on the topics related to the query.

Using the Shared-Data Overlay Network (SON), we have implemented a prototype of P2PShare that integrates three major DataRing services: ProbDB, a probabilistic database management service for relational data; WebSmatch, an environment for Web data integration; and P2Prec, a social-based P2P recommendation service for large-scale content sharing.

In [50], , we describe the demo of P2PShare's main services, e.g., gossiping topics of interest among friends, key- word querying for contents, and probabilistic queries over datasets.

6.2.3. View Selection in Distributed Data Warehousing

Participants: Zohra Bellahsène, Imen Mami.

Scientific data generate large amounts of data which have to be collected and stored for analytical purpose. One way to help managing and analyzing large amounts of data is data warehousing, whereby views over data are materialized [23]. At large scale, a data warehouse can be distributed. We have examined the problem of choosing a set of views and a set of data warehouse nodes at which these views should be materialized so that the full query workload is answered with the lowest cost. To address this problem, we extended our view selection method that we proposed for the centralized case. Thus, we modelled the distributed view selection problem as a Constraint Satisfaction Problem (CSP). Furthermore, we introduced the distributed AND-OR view graph, which can be seen as an extensive form of the AND-OR view graph to reflect the relation between views and communication network within the distributed scenario. The experiment results show that our approach provides better performance compared with the genetic algorithm in term of the solution quality (i.e., the quality of the obtained set of materialized views). We demonstrated experimentally that our approach provides better results in term of cost savings when the view selection is decided under space and maintenance cost constraints [44].

6.2.4. Scientific Workflow Management

Participants: Ayoub Ait Lahcen, Jonas Dias, Didier Parigot, Patrick Valduriez.

Scientific experiments based on computer simulations can be defined, executed and monitored using Scientific Workflow Management Systems (SWfMS). Several SWfMS are available, each with a different goal and a different engine. Due to the exploratory analysis, scientists need to run parameter sweep (PS) workflows, which are workflows that are invoked repeatedly using different input data. These workflows generate a large amount of tasks that are submitted to High Performance Computing (HPC) environments. Different execution models for a workflow may have significant differences in performance in HPC. However, selecting the best execution model for a given workflow is difficult due to the existence of many characteristics of the workflow that may affect the parallel execution.

In [36], we develop a study to show performance impacts of using different execution models in running PS workflows in HPC. Our study contributes by presenting a characterization of PS workflow patterns (the basis for many existing scientific workflows) and its behavior under different execution models in HPC. We evaluated four execution models to run workflows in parallel. Our study measures the performance behavior of small, large and complex workflows among the evaluated execution models. The results can be used as a guideline to select the best model for a given scientific workflow execution in HPC. Our evaluation may also serve as a basis for workflow designers to analyze the expected behavior of an HPC workflow engine based on the characteristics of PS workflows.

This work was done in the context of the the CNPq-Inria project DatLuge and FAPERJ-Inria P2Pcloud project .

In the context of SON, we also proposed a declarative workflow language based on service/activity rules. In [27], [46], we present a formal approach that combines component-based development with well-understood methods and techniques from the field of Attribute Grammars and Data-Flow Analysis in order to specify the behavior of P2P applications, and then construct an abstract representation (i.e., Data-Dependency Graph) to perform analyzes on it. This formal approach makes it possible to infer a dependency graph for SON applications that provides for automatic parallelization.

6.2.5. Plants identification and classification from social image data

Participants: Hervé Goëau, Alexis Joly, Saloua Litayem.

This work is done in collaboration with the botanists of the AMAP UMR team (CIRAD) and with Inria team IMEDIA. Inspired by citizen sciences, the main goal of this trans-disciplinary work is to speed up the collection and integration of raw botanical observation data, while providing to potential users an easy and efficient access to this botanical knowledge. We therefore did continue working intensively on plants identification and classification [54], [37], [38], [26]. We first developed a new interactive method [37] for the visual identification of plants from social image data. Contrary to previous content-based identification methods and systems that mainly relied on leaves, or in few other cases on flowers, it makes use of five different organs and plant's views including habit, flowers, fruits, leaves and bark. Thanks to an interactive query widget, the tagging process of the different organs and views is as simple as drag-and-drop operations and does not require any expertise in botany. All training pictures used by the system were continuously collected during one year through a crowdsourcing application and more than 17K images are now integrated. System-oriented and human-centered evaluations of the application show that the results are already satisfactory and therefore very promising in the long term to identify a richer flora.
Besides, we did continue working on leaf-based identification notably through the organization of and participation to ImageCLEF plant identification evaluation campaign 2012 [54].
Finally we did apply one of our former work related to multi-source shared-nearest neighbors clustering to an original experiment aimed at evaluating if we were able to automatically recover morphological classifications built by the botanists themselves [38]. The results are very promising, since all clusters discovered automatically could be easily matched to one node of a morphological tree built by botanists.

6.3. Scalable Data Analysis

6.3.1. StreamCloud

Participants: Vincenzo Gulisano, Patrick Valduriez.

Recent years have witnessed the growth of a new class of data-intensive applications that do not fit the DBMS query paradigm. Instead, the data arrive at high speeds taking the form of an unbounded sequence of values (data streams) and queries run continuously returning new results as new data arrive. Examples of data streams are sensor data (e.g. in environmental applications) or IP packets (e.g. in a network monitoring application). The unbounded nature of data streams makes it impossible to store the data entirely in bounded memory. Current research efforts have mainly focused on scaling in the number of queries and/or query operators having overlooked the scalability with respect to the stream volume.

Current Stream Processing Engines do not scale with the input load due to single-node bottlenecks. Additionally, they are based on static configurations that lead to either under or over-provisioning. In [21], [22], we present StreamCloud, a scalable and elastic stream processing engine for processing large data stream volumes. StreamCloud uses a novel parallelization technique that splits queries into subqueries that are allocated to independent sets of nodes in a way that minimizes the distribution overhead. Its elastic protocols exhibit low intrusiveness, enabling effective adjustment of resources to the incoming load. Elasticity is combined with dynamic load balancing to minimize the computational resources used. We present the system design, implementation and a thorough evaluation of the scalability and elasticity of the fully implemented system.

6.3.2. Mining Uncertain Data Streams

Participants: Reza Akbarinia, Florent Masseglia.

Dealing with uncertainty by using probabilistic approaches has gained increasing attention these past few years. One of the main requirements for uncertain data mining is the ability to discover Probabilistic Frequent Itemsets (PFI). However, PFI mining, particularly in uncertain data streams, is very challenging and needs the development of new techniques, since approaches designed for deterministic data are not applicable in this context. In [29], we propose an efficient solution for exact PFI mining over data streams with sliding windows. Our proposal includes efficient solutions for updating frequentness probability of itemsets and thus fast extraction of PFI, whenever transactions are added or removed from the sliding window. To the best of our knowledge, this is the first efficient solution for data stream PFI mining. We have conducted an extensive experimental evaluation of our approach over synthetic and real-world data sets; the results illustrate its very good performance.

6.3.3. Detecting Rare Events in Massive Datasets

Participant: Florent Masseglia.

In this work, we consider that rare events are very small clusters typically representing less than 0.01% of the entire dataset. Finding these abnormal events allows to identify the emergence of pos- sible anomalies in their very early stages. Such a scenario is generally difficult to handle as it lies at the frontier between outlier detection and clustering and is characterized by a clear challenge to avoid false nega- tives. To address this challenge, we take a backward approach and pro- pose RARE, a framework that identifies and isolates the abnormal/rare regions. The dense regions are identified using a radius-limited density- driven variant of k-means and adjacent regions are merged to form new regions. These newly formed regions are gradually augmented as long as a density-driven condition is respected. When no more dense regions are observed, the remaining data is clustered and presented for further analysis to human experts. The framework is tested on a medical appli- cation and compared against human analysis. The experiments show that rare events that were missed during human analysis because of the multivariate character of the data can be discovered by our approach.

This work is funded by the labex NUMEV and a patent application involving Inria, CNRS, UM2 and INSERM has been filled.

6.3.4. Highly Informative Feature Set Mining

Participant: Florent Masseglia.

For many textual collections, the number of features is often overly large. As these features can be very redundant, it is desirable to have a small, succinct, yet highly informative collection of features that describes the key characteristics of a dataset. Information theory is one such tool for us to obtain this feature collection. In [48], we mainly contribute to the improvement of efficiency for the process of selecting the most informative feature set over high-dimensional unlabeled data. We propose a heuristic theory for informative feature set selection from high dimensional data. Moreover, we design data structures that enable us to compute the entropies of the candidate feature sets efficiently. We also develop a simple pruning strategy that eliminates the hopeless candidates at each forward selection step. We test our method through experiments on real-world data sets, showing that our proposal is very efficient.

6.3.5. Clustering Users with Evolving Profiles in Usage Streams

Participant: Florent Masseglia.

Existing data stream models commonly assume that users' records or profiles in data streams will not be updated once they arrive. In many applications such as web usage, however, the users' records/profiles may evolve along time. This kind of streaming transactions are referred to as bi-streaming data (*i.e.* the data evolves temporally in two dimensions, the flowing of transactions as with the traditional data streams, and the evolving of users' profiles inside the streams, which makes bi-streaming data different from traditional data streams). The two-dimensional evolving of bi-streaming data brings difficulties on modeling and clustering for exploring the users' behaviors. In [49], we propose three models to summarize bi-streaming data, which are the batch model, the Evolving Objects (EO) model and the Dynamic Data Stream (DDS) model. Through creating, updating and deleting user profiles, the models summarize the behaviors of each user as an object. Based on these models, clustering algorithms are employed to identify the user groups. The proposed models are tested on a real-world data set showing that the DDS model can summarize the bi-streaming data efficiently and effectively, providing better basis for clustering user profiles than the other two models.

6.3.6. Scalable Mining of Small Visual Objects

Participants: Pierre Letessier, Julien Champ, Alexis Joly.

Automatically linking multimedia documents that contain one or several instances of the same visual object has many applications including: salient events detection, relevant patterns discovery in scientific data or simply web browsing through hyper-visual links. Whereas efficient methods now exist for searching rigid objects in large collections, discovering them from scratch is still challenging in terms of scalability, particularly when the targeted objects are rather small. In this work [40], we formally revisit the problem of mining or discovering such objects, and then generalized two kinds of existing methods for probing candidate object seeds: weighted adaptive sampling and hashing based methods. We then introduce a new hashing strategy, working first at the visual level, and then at the geometric level. Experiments conducted on millions of images show that our method outperforms state-of-the-art.

This method was integrated within a visual-based media event detection system in the scope of a French project called the transmedia observatory. It allows the automatic discovery of the most circulated images across the main news media (news websites, press agencies, TV news and newspapers). The main originality of the detection is to rely on the transmedia contextual information to denoise the raw visual detections and consequently focus on the most salient trans-media events. This work was presented at ACM Multimedia Grand Challenge 2012 [39]. The movie presented during this event is available at http://www.otmedia.fr/?p=217.