RESEARCH CENTER
Rennes - Bretagne-Atlantique

FIELD

Activity Report 2012

Section New Results

Edition: 2013-04-24

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.

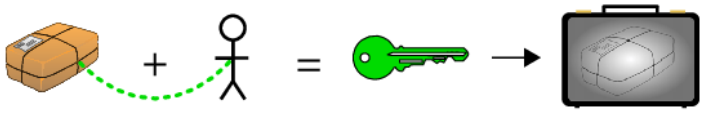


Figure 2. Key to a Ubi-Post briefcase

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

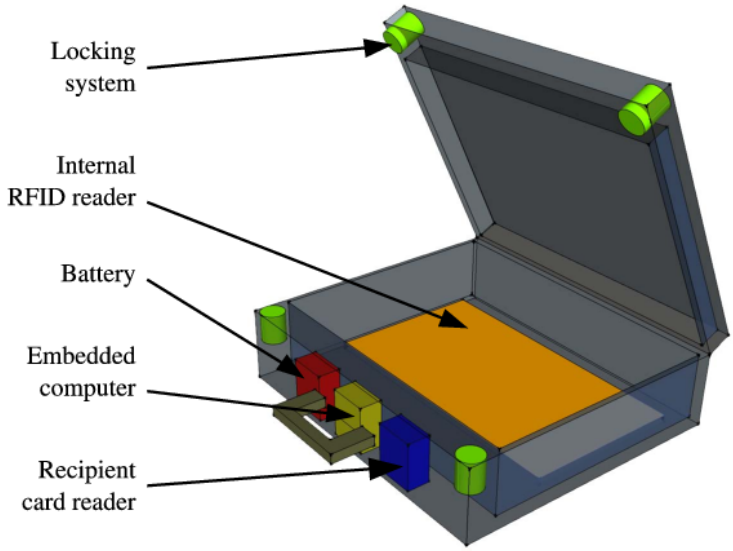


Figure 3. 3D view from the internal components of the Ubi-Post briefcase

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.

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.

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.

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.

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].

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.

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.

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 $O(nlog^3n)$, 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.

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).

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].

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]

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.

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).

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.

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+$N\left(N+1\right)/2$ unknowns, the six-dimensional electromagnetic field plus the $N\left(N+1\right)/2$ 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.

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.

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.

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.

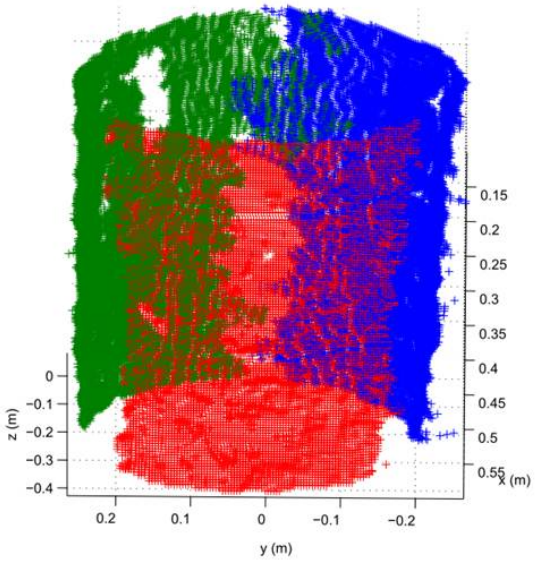


Figure 4. 3D silouhettes reconstructed with three depth-cameras - reconstructed points of a reference cylinder. Each color corresponds to one of the depth camera.

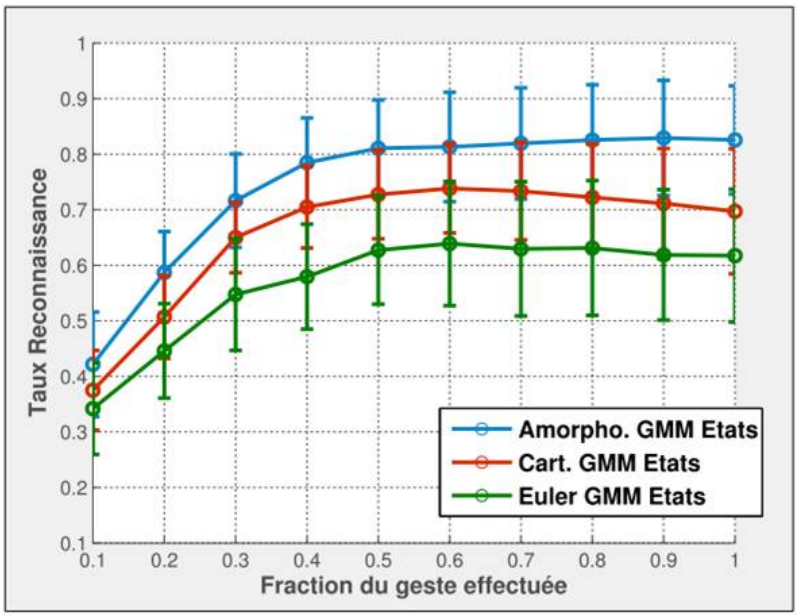


Figure 5. Early recognition of a motion performed by a new user with three different features: Cartesian, Euler and the proposed amorphological features.

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.

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].

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.

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).

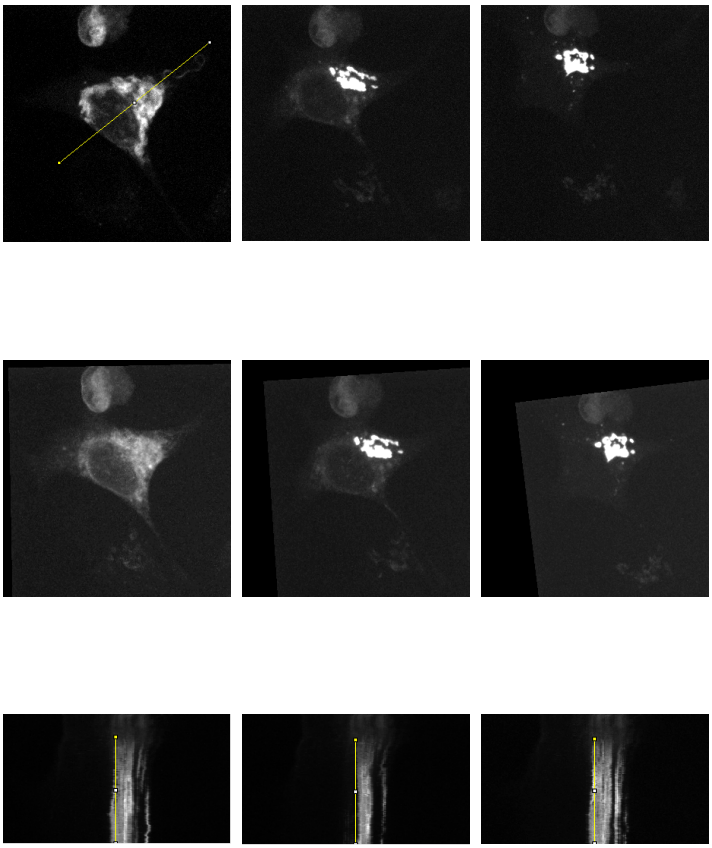


Figure 6. Real-time imaging of the synchronized trafficking of ManII-SBP-EGFP [21]. HeLa cells were transfected to express Ii-streptavidin as a hook and ManII-SBP-EGFP as a reporter. Release of the reporter was induced by addition of biotin and monitored using a spinning disk confocal microscope (F. 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).*

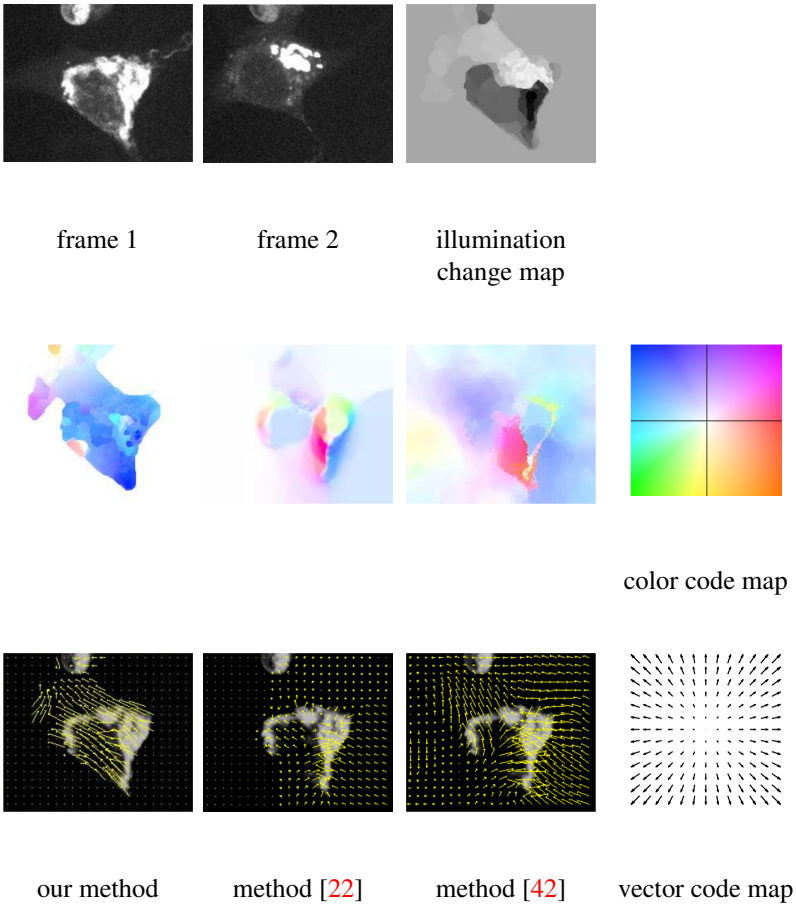


Figure 7. Comparison of our method with the methods of [22] and [42] on a sequence of "HeLa cells" (courtesy of F. Pérez UMR 144 CNRS Institut Curie, PICT-IBiSA).

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).

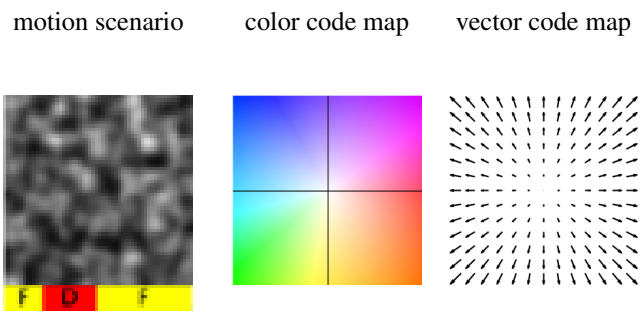
Partners: F. Waharte and J. Boulanger (UMR 144 CNRS PICT IBiSA Institut Curie)

6.6. Microtubules modeling for variational assimilation analysis

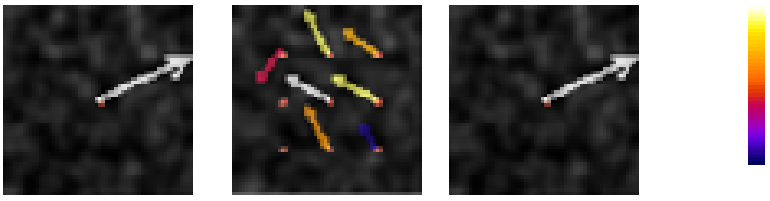
Participants: Pierre Allain, Charles Kervrann.

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.



STICS-based estimation of different flux and diffusion phases



Variational method with two different regularization parameters

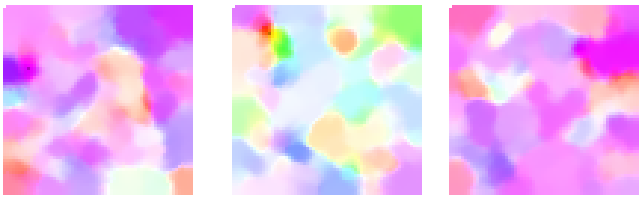


Figure 8. Analysis of STICS and variational methods on artificial image time series with three phases. First row: first frame of the sequence and temporal description of the 3 phases: F/Flux (i.e. directed flow), D/Diffusion (20 - 30 - 40 images) (left); coding maps of vector fields (middle and right). 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).

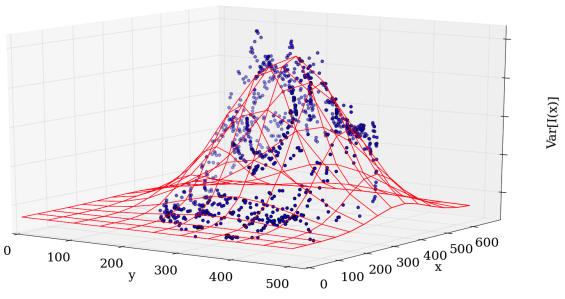


Figure 9. Variance prediction after noise model calibration using a fluorescein FD-FLIM reference stack acquired with a wide-field (WF) microscope (parametric model in red and measurements in blue).

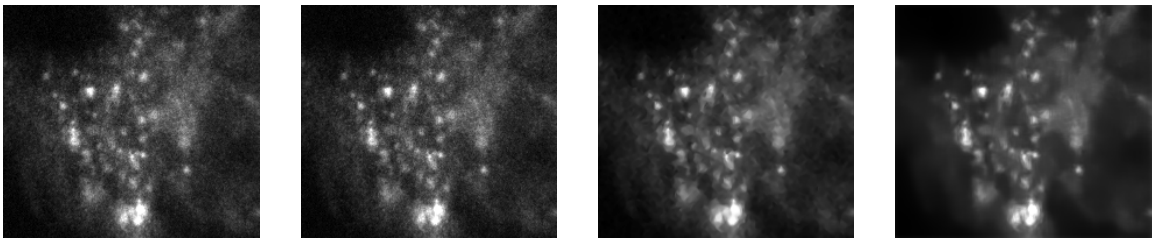


Figure 10. Denoising performance on a live cell image acquired in FD-FLIM (fluorescent tagged caveolin protein) using a confocal setup. From left to right: original image, results with BM3D *[27]* ND-SAFIR *[4] and our method.*

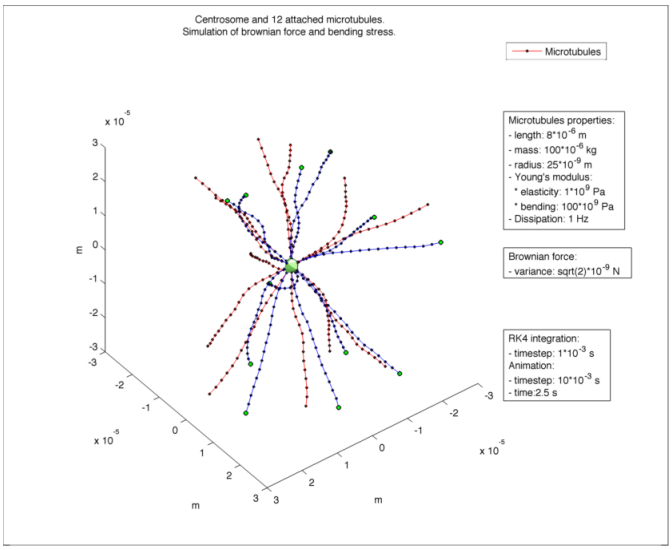


Figure 11. Simulation of growing microtubules in 3D.

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)

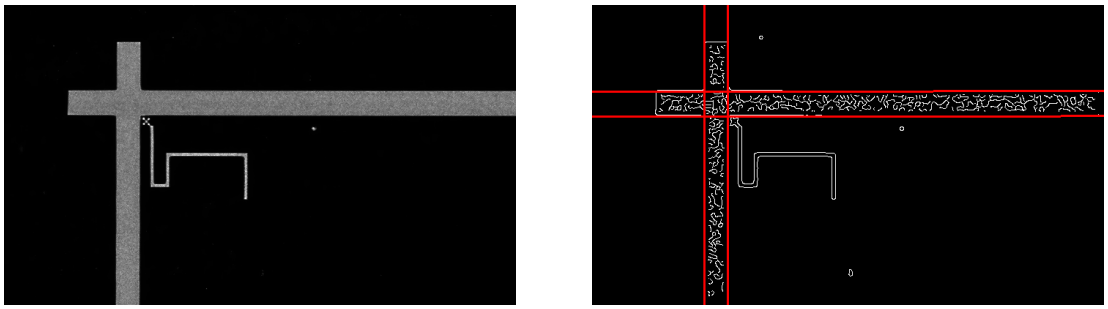


Figure 12. Detection of lines in an image (tool-slide) (pixel size: $3\mu m \times 3\mu m$*).*

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.

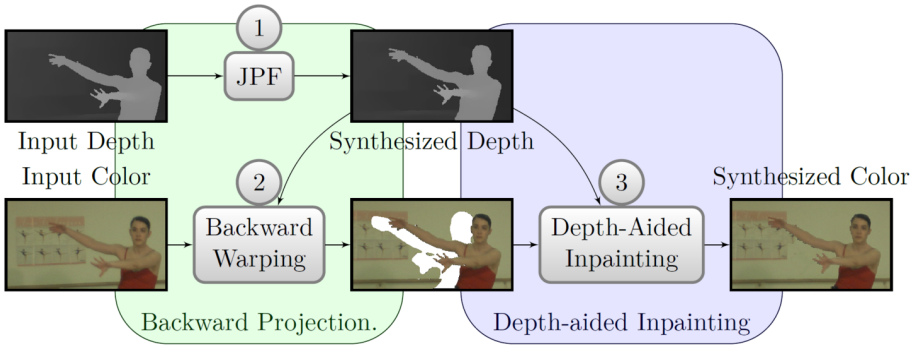


Figure 1. Virtual view generation chain, based on Joint Filling Projection. The depth map is jointly warped and inpainted. Depth-aided inpainting can be then used on disoccluded areas.

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.

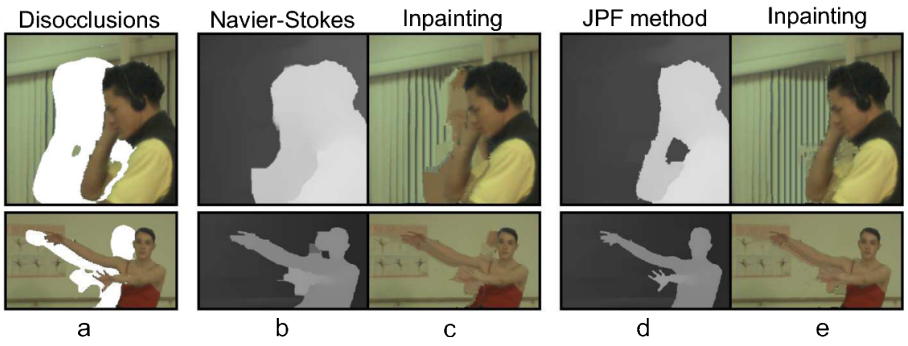


Figure 2. Disocclusion filling. (a) warped image before inpainting. Depth map inpainting: Navier-Stokes (b), JPF (d). Resulting depth-aided inpainting: Navier-Stokes (c), JPF (e).

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.

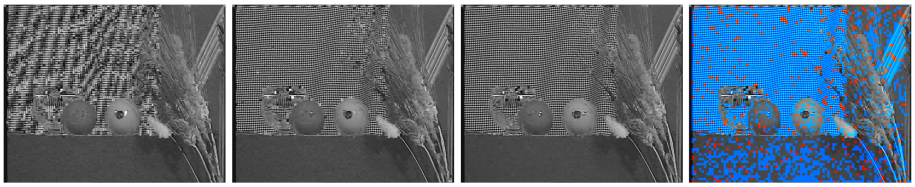


Figure 5. Spatial prediction for "Snook" with modes dynamically chosen according to a RD criterion with (a) H.264 Intra modes (High Profile), (b) LLE-based prediction, (c) Hybrid LLE-oMANE prediction and (d) selection rate of the two modes: LLE (red) and oMANE (blue).

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.

TASC Project-Team

6. New Results

6.1. Constraint and Abstract Interpretation

Participants: Marie Pelleau, Charlotte Truchet, Frédéric 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.

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).

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.

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 (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 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.

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.

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.

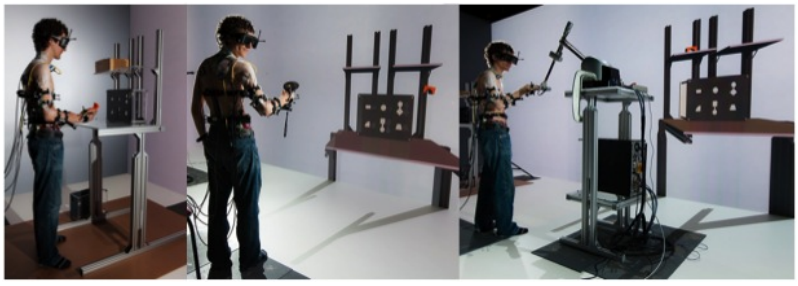


Figure 2. Simulation of an assembly task (Left in real, center in virtual, right in virtual with force-feedback)

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).

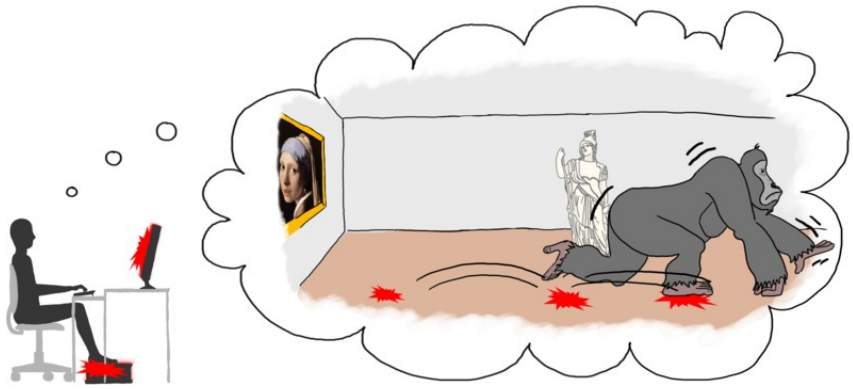


Figure 4. Concept of the King Kong Effects: Visual and Tactile vibrations inspired by special effects in movies enhance the sensation of walking in VE. Visual and Tactile feedbacks are generated at each step made in the VE.

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.

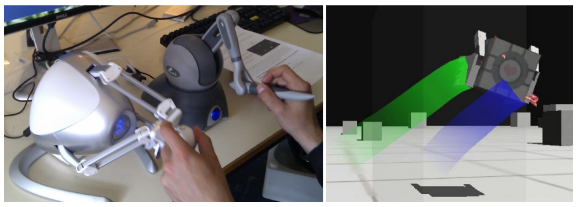


Figure 5. Bimanual pick-and-place task in a large virtual environment. A bimanual haptic setup made of two single-point devices (on left) allows to carry and displace a virtual cube using novel interactive techniques (on right).

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.

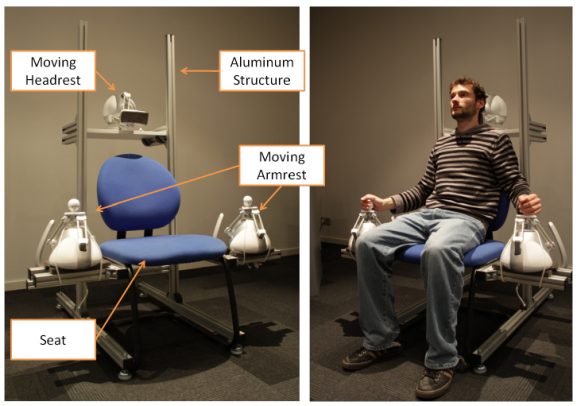


Figure 6. Prototype of the HapSeat. Left: seat structure with 3 force-feedback devices. Right: the system in use.

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.

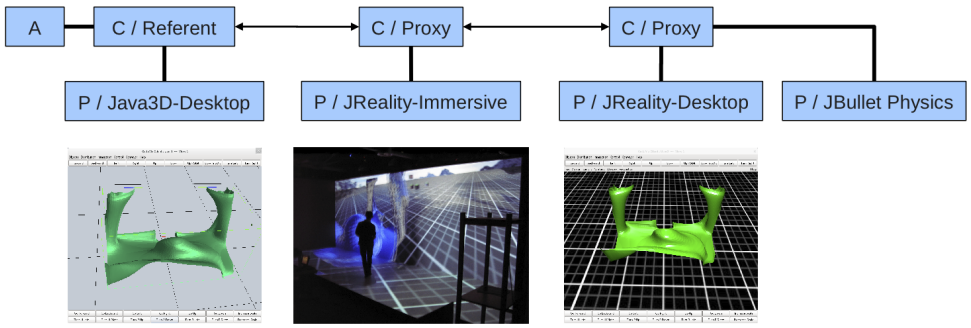


Figure 7. The PAC-C3D software architectural model makes interoperability possible between heterogeneous 3D viewers

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.

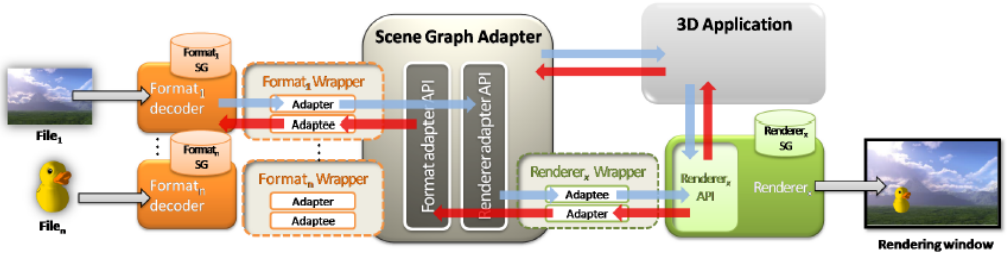


Figure 8. Our architecture allows the loading of any 3D graphics format simultaneously in any available rendering engine. The scene graph adapter is an interface that adapts a scene graph (SG) of a given format into a renderer scene graph and which also allows the rendering part to request this scene 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