RESEARCH CENTER
Paris - Rocquencourt

FIELD

Activity Report 2013

Section Application Domains

Edition: 2014-03-19

ABSTRACTION Project-Team

4. Application Domains

4.1. Certification of Safety Critical Software

Keywords: Absence of runtime error, Abstract interpretation, Certified compilation, Static analysis, Translation validation, Verifier.

Safety critical software may incur great damage in case of failure, such as human casualties or huge financial losses. These include many kinds of embedded software, such as fly-by-wire programs in aircrafts and other avionic applications, control systems for nuclear power plants, or navigation systems of satellite launchers. For instance, the failure of the first launch of Ariane 5 (flight Ariane 501) was due to overflows in arithmetic computations. This failure caused the loss of several satellites, worth up to $ 500 millions.

This development of safe and secure critical software requires formal methods so as to ensure that they do not go wrong, and will behave as specified. In particular, testing, bug finding methods, checking of models but not programs do not provide any guarantee that no failure will occur, even of a given type such as runtime errors; therefore, their scope is limited for certification purposes. For instance, testing can usually not be performed for *all* possible inputs due to feasibility and cost reasons, so that it does not prove anything about a large number of possible executions.

By contrast, program analysis methods such as abstract-interpretation-based static analysis are not subject to unsoundness, since they can *formally prove* the absence of bugs directly on the program, not on a model that might be erroneous. Yet, these techniques are generally incomplete since the absence of runtime errors is undecidable. Therefore, in practice, they are prone to false alarms (*i.e.*, they may fail to prove the absence of runtime errors for a program which is safe). The objective of certification is to ultimately eliminate all false alarms.

It should be noted that, due to the size of the critical codes (typically from 100 to 1000 kLOCs), only scalable methods can succeed (in particular, software model checking techniques are subject to state explosion issues). As a consequence, this domain requires efficient static analyses, where costly abstractions should be used only parsimoniously.

Furthermore, many families of critical software have similar features, such as the reliance on floating-point intensive computations for the implementation of control laws, including linear and non-linear control with feedback, interpolations, and other DSP algorithms. Since we stated that a proof of absence of runtime errors is required, very precise analyses are required, which should be able to yield no false alarm on wide families of critical applications. To achieve that goal, significant advantages can be found in the design of domain specific analyzers, such as ASTRÉE [30], [46], which has been initially designed specifically for synchronous embedded software.

Last, some specific critical software qualification procedures may require additional properties being proved. As an example, the DO-178 regulations (which apply to avionics software) require a tight, documented, and certified relation to be established between each development stage. In particular, compilation of high level programs into executable binaries should also be certified correct.

The ABSTRACTION project-team has been working on both proof of absence of runtime errors and certified compilation over the decade, using abstract interpretation techniques. Successful results have been achieved on industrial applications using the ASTRÉE analyzer. Following this success, ASTRÉE has been licensed to AbsInt Angewandte Informatik GmbH to be industrialized, and the ABSTRACTION project-team has strong plans to continue research on this topic.

4.2. Abstraction of Biological Cell Signaling Networks

Keywords: Biology, Health, Static analysis.

Protein-protein interactions consist in complexations and post translational modifications such as phosphorilation. These interactions enable biological organisms to receive, propagate, and integrate signals that are expressed as proteins concentrations in order to make decisions (on the choice between cell division and cell death for instance). Models of such interaction networks suffer from a combinatorial blow up in the number of species (number of non-isomorphic ways in which some proteins can be connected to each others). This large number of species makes the design and the analysis of these models a highly difficult task. Moreover the properties of interest are usually quantitative observations on stochastic or differential trajectories, which are difficult to compute or abstract.

Contextual graph-rewriting systems allow a concise description of these networks, which leads to a scalable method for modeling them. Then abstract interpretation allows the abstraction of these systems properties. First qualitative abstractions (such as over approximation of complexes that can be built) provide both debugging information in the design phases (of models) and static information that are necessary in order to make other computations (such as stochastic simulations) scale up. Then qualitative invariants also drive efficient quantitative abstractions (such as the reduction of ordinary differential semantics).

The work of the ABSTRACTION project-team on biological cell signaling networks ranges from qualitative abstractions to quantitative abstractions.

ALPAGE Project-Team

4. Application Domains

4.1. Overview

NLP tools and methods have many possible domains of application. Some of then are already mature enough to be commercialized. They can be roughly classified in three groups:

Human-computer interaction : mostly speech processing and text-to-speech, often in a dialogue context; today, commercial offers are limited to restricted domains (train tickets reservation...);

Language writing aid : spelling, grammatical and stylistic correctors for text editors, controlled-language writing aids (e.g., for technical documents), memory-based translation aid, foreign language learning tools, as well as vocal dictation;

Access to information : tools to enable a better access to information present in huge collections of texts (e.g., the Internet): automatic document classification, automatic document structuring, automatic summarizing, information acquisition and extraction, text mining, question-answering systems, as well as surface machine translation. Information access to speech archives through transcriptions is also an emerging field.

Experimental linguistics : tools to explore language in an objective way (this is related, but not limited to corpus linguistics).

Alpage focuses on applications included in the three last points, such as information extraction and (linguistic and extra-linguistic) knowledge acquisition (4.2), text mining (4.3), spelling correction (4.5) and experimental linguistics (4.6).

4.2. Information extraction and knowledge acquisition

Participants: Éric Villemonte de La Clergerie, Mickaël Morardo, Rosa Stern, Benoît Sagot.

The first domain of application for Alpage parsing systems is information extraction, and in particular knowledge acquisition, be it linguistic or not, and text mining.

Knowledge acquisition for a given restricted domain is something that has already been studied by some Alpage members for several years. Obviously, the progressive extension of Alpage parsing systems or even shallow processing chains to the semantic level increase the quality of the extracted information, as well as the scope of information that can be extracted. Such knowledge acquisition efforts bring solutions to current problems related to information access and take place into the emerging notion of *Semantic Web*. The transition from a web based on data (textual documents,...) to a web based on knowledge requires linguistic processing tools which are able to provide fine grained pieces of information, in particular by relying on high-quality deep parsing. For a given domain of knowledge (say, news wires or tourism), the extraction of a domain ontology that represents its key concepts and the relations between them is a crucial task, which has a lot in common with the extraction of linguistic information.

In the last years, such efforts have been targeted towards information extraction from news wires in collaboration with the Agence France-Presse (Rosa Stern was a CIFRE PhD student at Alpage and at AFP, and worked in 2013 within the ANR project EDyLex).

These applications in the domain of information extraction raise exciting challenges that require altogether ideas and tools coming from the domains of computational linguistics, machine learning and knowledge representation.

4.3. Processing answers to open-ended questions in surveys: vera

Participants: Benoît Sagot, Valérie Hanoka.

Verbatim Analysis is a startup co-created by Benoît Sagot from Alpage and Dimitri Tcherniak from Towers Watson, a world-wide leader in the domain of employee research (opinion mining among the employees of a company or organization). The aim of its first product, *vera*, is to provide an all-in-one environment for editing (i.e., normalizing the spelling and typography), understanding and classifying answers to open-ended questions, and relating them with closed-ended questions, so as to extract as much valuable information as possible from both types of questions. The editing part relies in part on SxPipe (see section 5.6) and Alexina morphological lexicons. Several other parts of *vera* have been co-developed by Verbatim Analysis and by Inria.

In 2013, Verbatim Analysis has bought Inria's part of the intellectual property of the first version of *vera*. A second version has been released, which is co-owned by Verbatim Analysis and Inria.

4.4. Multilingual terminologies and lexical resources for companies

Participants: Éric Villemonte de La Clergerie, Mickaël Morardo.

Lingua et Machina is a small company now headed by François Brown de Colstoun, a former Inria researcher, that provides services for developing specialized multilingual terminologies for its clients. It develops the WEB framework Libellex for validating such terminologies. A formal collaboration with ALPAGE has been set up, with the recruitment of Mikael Morardo in 2012 as an engineer, funded by Inria's DTI. He pursued his work on the extension of the web platform *Libellex* for the visualization and validation of new types of lexical resources. In particular, he has integrated a new interface for handling monolingual terminologies, lexical networks, and bilingual wordnet-like structures, including the WOLF.

4.5. Automatic and semi-automatic spelling correction in an industrial setting

Participants: Benoît Sagot, Kata Gábor, Éric Villemonte de La Clergerie.

NLP tools and resources used for spelling correction, such as large n-gram collections, POS taggers and finite-state machinery are now mature and precise. In industrial setting such as post-processing after large-scale OCR, these tools and resources should enable spelling correction tools to work on a much larger scale and with a much better precision than what can be found in different contexts with different constraints (e.g., in text editors). Moreover, such industrial contexts allow for a non-costly manual intervention, in case one is able to identify the most uncertain corrections. Alpage is working within the "Investissements d'avenir" project PACTE, headed by Numen, a company specialized in text digitalization, and three other partners. Kata Gábor is doing a post-doc funded by PACTE (see 6.7)

4.6. Experimental and quantitative linguistics

Participants: Benoît Crabbé, Margaret Grant, Juliette Thuilier, Benoît Sagot.

Alpage is a team that dedicates efforts in producing ressources and algorithms for processing large amounts of textual materials. These ressources can be applied not only for purely NLP purposes but also for linguistic purposes. Indeed, the specific needs of NLP applications led to the development of electronic linguistic resources (in particular lexica, annotated corpora, and treebanks) that are sufficiently large for carrying statistical analysis on linguistic issues. In the last 10 years, pioneering work has started to use these new data sources to the study of English grammar, leading to important new results in such areas as the study of syntactic preferences [51], [107], the existence of graded grammaticality judgments [67].

The reasons for getting interested for statistical modelling of language can be traced back by looking at the recent history of grammatical works in linguistics. In the 1980s and 1990s, theoretical grammarians have been mostly concerned with improving the conceptual underpinnings of their respective subfields, in particular through the construction and refinement of formal models. In syntax, the relative consensus on a generative-transformational approach [57] gave way on the one hand to more abstract characterizations of the language faculty [57], and on the other hand to the construction of detailed, formally explicit, and often implemented, alternative formulation of the generative approach [50], [76]. For French several grammars have been implemented in this trend, such as the tree adjoining grammars of [52], [59] among others. This general movement led to much improved descriptions and understanding of the conceptual underpinnings of both linguistic competence and language use. It was in large part catalyzed by a convergence of interests of logical, linguistic and computational approaches to grammatical phenomena.

However, starting in the 1990s, a growing portion of the community started being frustrated by the paucity and unreliability of the empirical evidence underlying their research. In syntax, data was generally collected impressionistically, either as ad-hoc small samples of language use, or as ill-understood and little-controlled grammaticality judgements (Schütze 1995). This shift towards quantitative methods is also a shift towards new scientific questions and new scientific fields. Using richly annotated data and statistical modelling, we address questions that could not be addressed by previous methodology in linguistics.

In this line, at Alpage we have started investigating the question of choice in French syntax with a statistical modelling methodology. In the perspective of better understanding which factors influence the relative ordering of post verbal complements across languages, Meg Grant (post-doc funded by the LabEx EFL), Juliette Thuilier (former PhD at Alpage), Anne Abeillé (LLF) and Benoit Crabbé designed psycholinguistic experiments (questionnaires and recall tasks) with a specific focus on French and on the influence of the animacy factor.

On the other hand we are also collaborating with the Laboratoire de Sciences Cognitives de Paris (LSCP/ENS) where we explore the design of algorithms towards the statistical modelling of language acquisition (phonological acquisition). This is currently supported by one PhD project.

In parallel, quantitative methods are applied to computational morphology, in collaboration with formal linguists from LLF (CNRS & U. Paris Diderot; Géraldine Walther, Olivier Bonami) and descriptive linguists from CRLAO (CNRS and Inalco; Guillaume Jacques) and HTL (CNRS, U. Paris Diderot and U. Sorbonne Nouvelle; Aimée Lahaussois) — see 6.5 .

ALPINES Team

4. Application Domains

4.1. Compositional multiphase Darcy flow in heterogeneous porous media

We study the simulation of compositional multiphase flow in porous media with different types of applications, and we focus in particular on reservoir/bassin modeling, and geological CO2 underground storage. All these simulations are linearized using Newton approach, and at each time step and each Newton step, a linear system needs to be solved, which is the most expensive part of the simulation. This application leads to some of the difficult problems to be solved by iterative methods. This is because the linear systems arising in multiphase porous media flow simulations cumulate many difficulties. These systems are non-symmetric, involve several unknowns of different nature per grid cell, display strong or very strong heterogeneities and anisotropies, and change during the simulation. Many researchers focus on these simulations, and many innovative techniques for solving linear systems have been introduced while studying these simulations, as for example the nested factorization [Appleyard and Cheshire, 1983, SPE Symposium on Reservoir Simulation].

4.2. Inverse problems

The research of F. Nataf on inverse problems is rather new since this activity was started from scratch in 2007. Since then, several papers were published in international journals and conference proceedings. All our numerical simulations were performed in FreeFem++.
We focus on methods related to time reversal techniques. Since the seminal paper by [M. Fink et al., Imaging through inhomogeneous media using time reversal mirrors. Ultrasonic Imaging, 13(2):199, 1991.], time reversal is a subject of very active research. The main idea is to take advantage of the reversibility of wave propagation phenomena such as it occurs in acoustics, elasticity or electromagnetism in a non-dissipative unknown medium to back-propagate signals to the sources that emitted them. Number of industrial applications have already been developped: touchscreen, medical imaging, non-destructive testing and underwater communications. The principle is to back-propagate signals to the sources that emitted them. The initial experiment, was to refocus, very precisely, a recorded signal after passing through a barrier consisting of randomly distributed metal rods. In [de Rosny and Fink. Overcoming the difraction limit in wave physics using a time-reversal mirror and a novel acoustic sink. Phys. Rev. Lett., 89 (12), 2002], the source that created the signal is time reversed in order to have a perfect time reversal experiment. Since then, numerous applications of this physical principle have been designed, see [Fink, Renversement du temps, ondes et innovation. Ed. Fayard, 2009] or for numerical experiments [Larmat et al., Time-reversal imaging of seismic sources and application to the great sumatra earthquake. Geophys. Res. Lett., 33, 2006] and references therein.

4.3. Numerical methods for wave propagation in multi-scale media

We are interested in the development of fast numerical methods for the simulation of electromagnetic waves in multi-scale situations where the geometry of the medium of propagation may be described through caracteristic lengths that are, in some places, much smaller than the average wavelength. In this context, we propose to develop numerical algorithms that rely on simplified models obtained by means of asymptotic analysis applied to the problem under consideration.

Here we focus on situations involving boundary layers and *localized* singular perturbation problems where wave propagation takes place in media whose geometry or material caracteristics are submitted to a small scale perturbation localized around a point, or a surface, or a line, but not distributed over a volumic sub-region of the propagation medium. Although a huge literature is already available for the study of localized singular perturbations and boundary layer pheneomena, very few works have proposed efficient numerical methods that rely on asymptotic modeling. This is due to their natural functional framework that naturally involves singular functions, which are difficult handle numerically. The aim of this part of our reasearch is to develop and analyze numerical methods for singular perturbation methods that are prone to high order numerical approximation, and robust with respect to the small parameter caracterizing the singular perturbation.

4.4. Data analysis in astrophysics

We focus on computationally intensive numerical algorithms arising in the data analysis of current and forthcoming Cosmic Microwave Background (CMB) experiments in astrophysics. This application is studied in collaboration with researchers from University Paris Diderot, and the objective is to make available the algorithms to the astrophysics community, so that they can be used in large experiments.

In CMB data analysis, astrophysicists produce and analyze multi-frequency 2D images of the universe when it was 5% of its current age. The new generation of the CMB experiments observes the sky with thousands of detectors over many years, producing overwhelmingly large and complex data sets, which nearly double every year therefore following the Moore's Law. Planck (http://www.rssd.esa.int/index.php?project=PLANCK) is a keystone satellite mission which has been developed under auspices of the European Space Agency (ESA). Planck has been surveying the sky since 2010, produces terabytes of data and requires 100 Petaflops per image analysis of the universe. It is predicted that future experiments will collect half petabyte of data, and will require 100 Exaflops per analysis as early as in 2020. This shows that data analysis in this area, as many other applications, will keep pushing the limit of available supercomputing power for the years to come.

ANGE Team

4. Application Domains

4.1. Fluids with complex rheology

Whereas the viscous effects can often be neglected in water flows, they have to be taken into account in situations such as avalanches, debris flows, pyroclastic flows, erosion processes,...*i.e.* when the fluid rheology becomes more complex. Gravity driven granular flows consist of solid particles commonly mixed with an interstitial lighter fluid (liquid or gas) that may interact with the grains and decrease the intensity of their contacts, thus reducing energy dissipation and favoring propagation. Examples include subaerial or subaqueous rock avalanches (*e.g.* landslides).

As mentioned above, the main issue is to propose models of reduced complexity, suitable for scientific computing and endowed with stability properties (continuous and/or discrete). In addition, models and their numerical approximations have to be confronted with experimental data, as analytical solutions are hardly accessible for these problems/models. A. Mangeney (IPGP) and N. Goutal (EDF) may provide useful data.

4.1.1. Arbitrary topography

Most shallow water type models are derived under the assumption of small/ smooth bottom variations whereas in practice the topography along which the flow (avalanche, debris flow,...) occurs can be quite steep and rough. An improved Saint-Venant system, due to Savage-Hutter, and valid for large slopes and small slope variations has been proposed. A new model relaxing all restrictions upon the topography has been proposed for shallow water flows by Bouchut *et al.* [24], [27]. The extension of this work to the case of models with distributed velocities along the vertical axis is an important objective with many applications (landslides, avalanches,...).

4.1.2. Erosion and sedimentation

The sediment transport modelling is of major interest in terms of applications. It also raises interesting issues from a numerical aspect. This is an example of coupling between the flow and another phenomenon, namely the deformation of the bottom of the basin that can be carried out either by bed load where the sediment has its own velocity or suspended load in which the particles are mostly driven by the flow. This phenomenon involves different time scales and nonlinear retroactions; hence the need for accurate mechanical models and very robust numerical methods. In collaboration with industrial partners (EDF–LNHE), the team already works on the improvement of numerical methods for existing (mostly empirical) models but our aim is also to propose new (quite) simple models that contain important features and satisfy some basic mechanical requirements. The extension of our 3D models to the transport of weighted particles can also be here of great interest.

4.2. Ecology and sustainable energies

Sustainable development and environment preservation have a growing importance and scientists have to address difficult issues such as: management of water resources, renewable energy production, biogeochemistry of oceans, resilience of society w.r.t. hazardous flows,...

4.2.1. Hydrodynamics-biology coupling

Nowadays, simulations of the hydrodynamic regime of a river, a lake or an estuary, are not restricted to the determination of the water depth and the fluid velocity. They have to predict the distribution and evolution of external quantities such as pollutants, biological species or sediment concentration.

4.2.1.1. Hydrodynamics-biology coupling for algae culture and biofuel production

The potential of micro-algae as a source of biofuel and as a technological solution for CO_2 fixation is the subject of intense academic and industrial research. Large-scale production of micro-algae has potential for biofuel applications owing to the high productivity that can be attained in high-rate raceway ponds.

One of the key challenges in the production of micro-algae is to maximize algae growth with respect to the exogenous energy that must be used (paddlewheel, pumps,...). There is a large number of parameters that need to be optimized (characteristics of the biological species, raceway shape, stirring provided by the paddlewheel); consequently our strategy is to develop efficient models and numerical tools to reproduce the flow induced by the paddlewheel and the evolution of the biological species within this flow. Here, mathematical models can greatly help us reduce experimental costs.

Owing to the high heterogeneity of raceways due to gradients of temperature, light intensity and nutrient availability through water height, we cannot use depth-averaged models. We adopt instead more accurate multilayer models that have recently been proposed.

It is clear however that many complex physical phenomena have to be added to our model, such as the effect of sunlight on water temperature/ density, evaporation and external forcing (wind).

4.2.1.2. Lacustrian ecosystems

Many problems previously mentioned also arise in larger scale systems like lakes. Hydrodynamics of lakes is mainly governed by atmospheric forcing terms: wind, temperature variations,...

If the interactions between hydrodynamics and biology are known via laboratory experiments, it is more difficult to predict the evolution – especially for the biological quantities – in a real and heterogeneous system. The objective is to model and reproduce the hydrodynamics modifications due to forcing term variations (in time and space). We are typically interested in phenomena such as eutrophication, development of harmful bacteria (cyanobacteria) and upwelling phenomena.

4.2.2. Marine energies

One of the booming lines of business is the field of renewable and decarbonated energies. In particular in the marine realm, several processes have been proposed in order to produce electricity thanks to the recovering of wave, tidal and current energies. We may mention water-turbines, buoys turning variations of the water height into electricity or turbines motioned by currents. Although these processes produce an amount of energy which is less substantial than in thermal or nuclear power plants, they have smaller dimensions and can be set up more easily.

The fluid energy has a kinetic and potential part. The buoys use the potential energy whereas the turbines are activated by currents. To become economically relevant, these systems need to be optimized (shape, position, durability, ...) in order to improve their productivity. This is a complex and original issue which requires efficient numerical tools.

Some processes are currently running. However, they have not been studied from an optimization point of view. While for the construction of a harbour, the goal is to minimize swell, in our framework we intend to maximize the wave energy. A key-point is the optimization of the bathymetry in a given geometrical domain which influences the swell and thus the effectiveness of processes. Optimization involving fluid mechanics is quite complex. Although such an approach seems innovative, it clearly requires the development of methodological tools. In a second step, experiments will be necessary for the validation.

AOSTE Project-Team

4. Application Domains

4.1. Multicore System-on-Chip design

Synchronous formalisms and GALS or multiclock extensions are natural model representations of hardware circuits at various abstraction levels. They may compete with HDLs (Hardware Description Languages) at RTL and even TLM levels. The main originality of languages built upon these models is to be based on formal *synthesis* semantics, rather than mere simulation forms.

The flexibility in formal Models of Computation and Communication allows specification of modular Latency-Insensitive Designs, where the interconnect structure is built up and optimized around existing IP components, respecting some mandatory computation and communication latencies prescribed by the system architect. This allows a real platform view development, with component reuse and timing-closure analysis. The design and optimization of interconnect fabric around IP blocks transform at modeling level an (untimed) asynchronous versions into a (scheduled) multiclock timed one.

Also, Network on Chip (NoC) design may call for computable switching patterns, just like computable scheduling patterns were used in (predictable) Latency-Insensitive Design. Here again formal models, such as Cyclo-static dataflow graphs and extended Kahn networks with explicit routing schemes, are modeling elements of choice for a real synthesis/optimization approach to the design of systems. New parallel architecture paradigms, such as GPU co-processors or Massively Parallel Processor Arrays (MPPA) form natural targets as NoC-based platforms.

Multicore embedded architecture platform may be represented as Marte UML component diagrams. The semantics of concurrent applications may also be represented as Marte behavior diagrams embodying precise MoCCs. Optimized compilations/syntheses rely on specific algorithms, and are represented as model transformations and allocation (of application onto architecture).

Our current work aims thus primarily at providing Theoretical Computer Science foundations to this domain of multicore embedded SoCs, with possibly efficient application in modeling, analysis and compilation wherever possible due to some natural assumptions. We also deal with a comparative view of Esterel and SystemC TLM for more practical modeling, and the relation between the Spirit IP-Xact interface standard in SoC domain with its Marte counterpart.

4.2. Automotive and avionic embedded systems

Model-Driven Engineering is in general well accepted in the transportation domains, where design of digital software and electronic parts in usually tighly coupled with larger aspects of system design, where models from physics are being used already. The formalisms AADL (for avionics) and AutoSar [66] (for automotive) are providing support for this, unfortunately not always with a clean and formal semantics. Thus there is a strong need here for approaches that bring closer together formal methods and tools on the one hand, engineering best practices on the other hand.

From a structural point of view AUTOSAR succeeded in establishing a framework that provides significant confidence in the proper integration of software components from a variety of distinct suppliers. But beyond those structural (interface) aspects, dynamic and temporal views are becoming more of a concern, so that AUTOSAR has introduced the AUTOSAR Specification of Timing Extension. AUTOSAR (discrete) timing models consist of timing descriptions, expressed by events and event chains, and timing constraints that are imposed on these events and event chains.

An important issue in all such formalisms is to mix in a single design framework heterogeneous time models and tasks: based on different timebases, with different triggering policy (event-triggered and time-triggered), and periodic and/or aperiodic tasks, with distinct periodicity if ever. Adequate modeling is a prerequisite to the process of scheduling and allocating such tasks onto complex embedded architectural platforms (see AAA approach in foundation section 3.3). Only then can one devise powerful synthesis/analysis/verification techniques to guide designers towards optimized solutions.

Traceability is also an important concern, to close the gap between early requirements and constraints modelling on the one hand, verification and correct implementation of these constraints at the different levels of the development on the other hand.

ARAMIS Team

4. Application Domains

4.1. Introduction

We develop different applications of our new methodologies to brain pathologies, mainly neurodegenerative diseases, epilepsy and cerebrovascular disorders. These applications aim at:

- better understanding the pathophysiology of brain disorders;
- designing biomarkers of pathologies for diagnosis, prognosis and assessment of drug efficacy;
- developping brain computer interfaces for clinical applications.

These applications are developed in close collaboration with biomedical researchers of the ICM and clinicians of the Pitié-Salpêtrière hospital.

4.2. Understanding brain disorders

The approaches that we develop allow to characterize anatomical and functional alterations, thus making it possible to study these alterations in different clinical populations. This can provide provide new insights into the mechanisms and progression of brain diseases. This typically involves the acquisition of neuroimaging data in a group of patients with a given pathology and in a group of healthy controls. Measures of anatomical and functional alterations are then extracted in each subject (for instance using segmentation of anatomical structures, shape models or graph-theoretic measures of functional connectivity). Statistical analyses are then performed to identify: i) significant differences between groups, ii) correlations between anatomical/functional alterations on the one hand, and clinical, cognitive or biological measures on the other hand, iii) progression of alterations over time.

We propose to apply our methodologies to study the pathophysiology of neurodegenerative diseases (mostly Alzheimer's disease and fronto-temporal dementia), epilepsy, cerebrovascular pathologies and neurodevelopmental disorders (Gilles de la Tourette syndrome). In neurodegenerative diseases, we aim at establishing the progression of alterations, starting from the early and even asymptomatic phases. In Gilles de la Tourette syndrome, we study the atypical anatomical patterns that may contribute to the emergence of symptoms. In epilepsy, we aim at studying the relationships between the different functional and structural components of epileptogenic networks.

4.3. Biomarkers for diagnosis, prognosis and clinical trials

Currently, the routine diagnosis of neurological disorders is mainly based on clinical examinations. This is also true for clinical trials, aiming to assess the efficacy of new treatments. However, clinical diagnoses only partially overlap with pathological processes. For instance, the sensitivity and specificity of clinical diagnosis of Alzheimer's disease (AD) based on established consensus criteria are of only about 70-80% compared to histopathological confirmation. Furthermore, the pathological processes often begin years before the clinical symptoms. Finally, clinical measures embed subjective aspects and have a limited reproducibility and are thus not ideal to track disease progression. It is thus crucial to supplement clinical examinations with biomarkers that can detect and track the progression of pathological processes in the living patient. This has potentially very important implications for the development of new treatments as it would help: i) identifying patients with a given pathology at the earliest stage of the disease, for inclusion in clinical trials; ii) providing measures to monitor the efficacy of treatments.

The derivation of biomarkers from image analysis approaches requires large-scale validation in well-characterized clinical populations. The ARAMIS team is strongly engaged in such efforts, in particular in the field of neurodegenerative disorders. To that purpose, we collaborate to several national studies (see section Partnerships) that involve multicenter and longitudinal acquisitions. Moreover, ARAMIS is strongly involved in the CATI which manages over 15 multicenter studies, including the national cohort MEMENTO (2000 patients).

4.4. Brain computer interfaces for clinical applications

A brain computer interface (BCI) is a device aiming to decode brain activity, thus creating an alternate communication channel between a person and the external environment. BCI systems can be categorized on the base of the classification of an induced or evoked brain activity. The central tenet of a BCI is the capability to distinguish different patterns of brain activity, each being associated to a particular intention or mental task. Hence adaptation, as well as learning, is a key component of a BCI because users must learn to modulate their brainwaves to generate distinct brain patterns. Usually, a BCI is considered a technology for people to substitute some lost functions. However, a BCI could also help in clinical rehabilitation to recover motor functions. Indeed, in current neuroscience-based rehabilitation it is recognized that protocols based on mental rehearsal of movements (like motor imagery practicing) are a way to access the motor system because they can induce an activation of sensorimotor networks that were affected by lesions. Hence, a BCI based on movement imagery can objectively monitor patient's progress and their compliance with the protocol, monitoring that they are actually imagining movements. It also follows that feedback from such a BCI can provide patients with an early reinforcement in the critical phase when there is not yet an overt sign of movement recovery. The BCI approaches that we develop are based on the characterization of the information contained in the functional connectivity patterns. We expect to significantly increase the performance of the BCI system with respect to the sole use of standard power spectra of the activity generated by single local brain areas. Such an improvement will concretely provide the user with a more precise control of the external environment in open-loop BCI tasks and a more coherent feedback in the closed-loop BCI schemes.

ARLES Project-Team

4. Application Domains

4.1. Pervasive Software Applications

The ARLES project-team is interested in the application of pervasive computing, and as such considers various application domains, especially considering the increasing pervasiveness of the digital world. However, we examine exploitation of our results for specific applications, as part of the experiments that we undertake to validate our research results through prototype implementation. Applications that we consider in particular include demonstrators developed in the context of the European and National projects to which we contribute (§ 7).

4. Application Domains

4.1. Panorama: Living Labs, Smart Cities

AxIS addresses any applicative field which has the following features:

a) requiring usage/data storage, preprocessing and analysis tools

- for designing, evaluating and improving huge evolving hypermedia information systems (mainly Web-based ISs), for which end-users are of primary concern,
- for a better understanding of the usage of a service/product via data mining techniques and knowledge management,
- for social network analysis (for example in Web 2.0 applications, Business Intelligence, Sustainable Development, etc.).

b) requiring user-driven innovation methods.

Even if our know-how, methods and algorithms have a cross domain applicability, our team chooses to focus on **Living Lab projects** (and mainly related to **Sustainable Development for Smart Cities**) [13], [12] which imply user involvement for the generation of future services/products. Indeed, following the Rio Conference (1992) and the Agenda for the 21st Century, local territories are now directly concerned with the set up of actions for a sustainable development. In this frame, ICT tools are supposed to be very efficient to re-engage people in the democratic process and to make decision-making more transparent, inclusive and accessible. So, sustainable development is closely associated with citizen participation. The emerging research field of e-democracy (so called Digital Democracy or eParticipation), concerned with the use of communications technologies such as the Internet to enhance the democratic processes is now a very active field. Though still in its infancy, a lot of literature is already available (see for instance: http://itc.napier.ac.uk/ITC/publications.asp for a global view of work in Europe) and numerous different topics are addressed in the field.

Our experience particularly stressed on the following applicative domains:

- Transportation systems & Mobility (cf. Section 4.2),
- Tourism (cf. Section 4.3),
- User Involvement in Silver Economy, Environment, Energy and e-government (cf. Section 4.4).

4.2. Transportation Systems & Mobility

Major recent evolutions in Intelligent Transportation Systems (ITS) are linked to rapid changes in communication technologies, such as ubiquitous computing, semantic web, contextual design. A strong emphasis is now put on mobility improvements. In addition to development of sustainable transportation systems (better ecological vehicles' performance, reduction of impacts on town planning etc.) these improvements concern also mobility management, that is specific measures to encourage people to adopt new mobility behaviour such as public transportation services rather than their personal car. These prompting measures concern for instance the quality of traveller's information systems for trip planning, the ability to provide real time recommendations for changing transportation means according to traffic information, and the quality of embedded services in vehicles to provide enhanced navigation aids with contextualised and personalised information.

Since 2004, AxIS has been concerned with mobility projects :

- PREDIT (2004-2007): The MobiVIP project has been an opportunity to collaborate with local Institutions ("Communauté d'Agglomération de Sophia Antipolis - CASA") and SMEs (VU Log) and to apply AxIS' know-how in data and web mining to the field of transportation systems.
- Traveller's information systems and recommender systems have been studied with the evaluation of two CASA web sites : the "Envibus" web site which provides information about a bus network and the "Otto&co" web site support car-sharing.
- Advanced transportation systems has been studied in PREDIT TIC TAC (2010-2012): this project aimed at optimizing travel time by providing in an aera with weak transportation services, a just in time on demand shuttle based on real time information. It was for AxIS the opportunity to experiment user implication in the design of a new travel information system called MOBILTIC.
- User Experience: in the ELLIOT project (2011-2013), the Mobility scenario is addressed in relation to information on air quality and noise and the use of Internet of Things (IoT).

4.3. Tourism

As tourism is a highly competitive domain, local tourism authorities have developed Web sites in order to offer of services to tourists. Unfortunately, the way information is organised does not necessarily meet Internet users expectations and numerous improvements are necessary to enhance their understanding of visited sites. Thus, even if only for economical reasons, the quality and the diversity of tourism packages have to be improved, for example by highlighting cultural heritage.

Again to illustrate our role in such a domain, Let us cite some past projects where AxIS is involved related mainly to **Semantic Web Mining** [3]. In our case, a) we exploit ontologies and semantic data for improving usage analysis, personalised services, the quality of results of search engines and for checking the content of an IS and also b) we exploit usage data for updating ontologies.) and Information Retrieval.

- Research has been carried out using log files from the city of Metz. This city was chosen because its Web site is in constant development and has been awarded several times, notably in 2003, 2004 and 2005 in the context of the Internet City label. The objective was to extract information about tourists behaviours from this site log files and to identify possible benefits in designing or updating a tourism ontology.
- Providing Tourism Information linked to Transportation information: AxIS has already studied recommender systems in order to provide users with personalised transportation information while looking for tourism information such as cultural information, leisure etc. (cf. our recommender Be-TRIP (2006) based on CBR*Tools).
- In the context of HOTEL-REF-PACA project, we aimed to better refer the web sites of hotels/campings from the region of TOURVAL in PACA (mainly Vésubie territory), with an approach based on a better understanding of usage from the internauts. To address this, we proposed and adopted a multidisciplinary approach combining various AxIS know-how: knowledge engineering (ontology in tourism), data mining (analysis of Google logs, hotel web site logs and user queries, visual behaviours from eye tracker), Ergonomics (clustering of hotel web sites based on their ergonomical quality).
- Several contacts (PACA, France Living Labs, Island of the Reunion) have been done related to projects in tourism and eco-tourism.

[3] By Semantic Web Mining, we mean the mutual benefits between two communities Semantic Web and Web Mining

4.4. User Involvement in Silver Economy, Environment, Energy and E-governement

Below are some topics where AxIS was or is involved in:

- **Preprocessing and analysing collective usage data and social networks** from group discussions related to design process: see ANR Intermed (2009) and FP7 Elliot where citizen generate ideas in terms of specific environmental sensors based services according to their needs.
- **Methods and tools for supporting open innovation based on open data**: a first work was made in 2010 with the CDISOD Color action related Public Data in collaboration with Fing (Marseille) and ADEME (Sophia Antipolis). We pursue such a study in the context of FP7 Elliot by providing to citizen environmental data (air quality and noise) issued from citizen and/or territories sensors.

All AxIS topics are relevant for these domains. Let us cite: social network analysis, personalization and information retrieval, recommender systems, expert search, design and evaluation of methods and tools for open innovation and user co-creation in the context of Living Labs, usage mining, mining data streams.

We have addressed specific works:

- **Silver Economy - Health & Well Being**: Axis contributed in 2010-2011 to a Living Lab characterisation in Health domain, study conducted by R. Picard (CGIET [4] via the participation of a working group (M. Pallot) and the visit of several European Living Labs, which operate in the domain of Health and Autonomy. B. Trousse as Inria representative of ICT usage lab involved in Health and Autonomy was also interviewed. This year Axis team managed the Green Services use case in the context of the achieved FP7 ELLIOT project involving pollution citizen sensors and in relation to health and Well being (targeted users with respiratory problems). interviews. This use case has been evaluated as "Good practice" by the international *Design for All* foundation (Awards 2014). Two ANR proposals involving France Living Labs and/or our living lab have been deposit with "Cité du Design" and University of Lorraine (cf. Sections 7.2.5 and 7.2.4). Let us note that France Living Labs is involved in the Silver Economy contract (cf. Section 7.2.6).
- **Energy**: the main AxIS topic here was usage analysis in the context of an energy challenge in an enterprise (ECOFFICES) taking into account the complex and real situation (installation for more than 400 sensors, differences between the three concerned teams, differences between the offices). Such an analysis aims to correlate team/office energy consuming, team/office eco-responsible behaviours and team/office profile. In 2012, our team was involved in a second project ECOFAMILIES aiming to co-design with families user interfaces for energy monitoring.
- **E-government**: The future Internet will bring a growing number of networked applications (services), devices and individual data (including private ones) to end-users. The important challenges are the organization of their access, and the guarantee of trust and privacy. The objectives of the PIMI [5] project (cf. section 7.2.1) are the definition of a design environment and a deployment platform for Personal Information Management system (PIM). The future PIM must provide the end-user personal data access with services that are relevant to his needs. In order to take mobility into account, the PIM will be accessed both by mobile devices (smart-phones) and Personal Computers. With the increasing number of services and associated data being accessible through Internet, the number and complexity of PIM will augment dramatically in the near future. This will require strong research investment in a number of topics, all contributing to the expected usability and accessibility of Individual Information Spaces for the end-user.

[4] CGIET: ”Conseil Général de l’Economie, de l’Industrie, de l’Energie et des Technologies”

[5] Personal Information Management through Internet

BANG Project-Team

4. Application Domains

4.1. Proliferation dynamics and its control

This domain of research has historically been - and is still - very active in the Bang team, which is reflected in particular in B. Perthame's book of 2007 “Transport equations in biology” [1]. It may presently be divided in:

- Cell division cycle in structured cell populations.
- Physiological and pharmacological control of cell proliferation.
- Optimisation of cancer chemotherapy and cancer chemotherapy.
- Protein polymerisation and application to amyloid diseases.
- Inverse problem for growth-fragmentation equations.

4.2. Tissue growth, regeneration and cell movements

This research activity aims at studying mathematical models related to tumour development and tissue organisation. Among the many biological aspects, examples are:

- Biomedical aspects of cell-cell interactions at the local and whole organ level.
- Migration of cells in tissues.
- Growth control of living tissues and organs.
- Regenerative medicine.
- Early embryology, and biomechanical aspects of cell interactions.
- Chemotaxis, self-organisation in cell populations.

4.3. Neurosciences

Cortical networks are constituted of a large number of statistically similar neurons in interaction. Each neuron has a nonlinear dynamics and is subject to noise. Moreover, neurological treatment involve several timescales. Multiscale analysis, both in spatial (number of cells) and temporal hence also constitute mathematical foundations of our approaches to neurosciences. In addition to the techniques described in section 3.1 - 3.4, our approach of the activity of large cortical areas involve:

- limit theorems of stochastic interacting particles systems, such as coupling methods or large deviations techniques, as used in mathematical approaches to the statistical physics of gases
- bifurcation analysis of deterministic and stochastic differential equations used to analyse the qualitative behaviour of networks
- singular perturbation theory, geometrical and topological approaches in dynamical systems used to uncover the dynamics in the presence of multiple timescales.

4.4. Geophysical flows and environment

The BANG team has split in December 2012, giving rise to another team, ANGE (https://team.inria.fr/ange/), specialised in complex geophysical flows in interaction with the environment. Free surface flows as tsunamis, flows in rivers and costal areas and their ecological consequences are typical examples of applications developed in this new Inria team, based on algorithms for the free-surface Navier-Stokes equations.

CAD Team

4. Application Domains

4.1. Domain

Computer Aided Design and Computer Graphics are two Application Domains.

CASCADE Project-Team

4. Application Domains

4.1. Privacy for the Cloud

Many companies have already started the migration to the Cloud and many individuals share their personal informations on social networks. While some of the data are public information, many of them are personal and even quite sensitive. Unfortunately, the current access mode is purely right-based: the provider first authenticates the client, and grants him access, or not, according to his rights in the access-control list. Therefore, the provider itself not only has total access to the data, but also knows which data are accessed, by whom, and how: privacy, which includes secrecy of data (confidentiality), identities (anonymity), and requests (obliviousness), should be enforced. Moreover, while high availability can easily be controlled, and thus any defect can immediately be detected, failures in privacy protection can remain hidden for a long time. The industry of the Cloud introduces a new implicit trust requirement: nobody has any idea at all of where and how his data are stored and manipulated, but everybody should blindly trust the providers. The providers will definitely do their best, but this is not enough. Privacy-compliant procedures cannot be left to the responsibility of the provider: however strong the trustfulness of the provider may be, any system or human vulnerability can be exploited against privacy. This presents too huge a threat to tolerate. *The distribution of the data and the secrecy of the actions must be given back to the users. It requires promoting privacy as a global security notion.*

In order to protect the data, one needs to encrypt it. Unfortunately, traditional encryption systems are inadequate for most applications involving big, complex data. Recall that in traditional public key encryption, a party encrypts data to a single known user, which lacks the expressiveness needed for more advanced data sharing. In enterprise settings, a party will want to share data with groups of users based on their credentials. Similarly, individuals want to selectively grant access to their personal data on social networks as well as documents and spreadsheets on Google Docs. Moreover, the access policy may even refer to users who do not exist in the system at the time the data is encrypted. Solving this problem requires an entirely new way of encrypting data.

A first natural approach would be **fully homomorphic encryption** (FHE, see above), but a second one is also **functional encryption**, that is an emerging paradigm for public-key encryption: it enables more fine-grained access control to encrypted data, for instance, the ability to specify a decryption policy in the ciphertext so that only individuals who satisfy the policy can decrypt, or the ability to associate keywords to a secret key so that it can only decrypt documents containing the keyword. Our work on functional encryption centers around two goals:

1. to obtain more efficient pairings-based functional encryption;
2. and to realize new functionalities and more expressive functional encryption schemes.

Another approach is **secure multi-party computation protocols**, where interactivity might provide privacy in a more efficient way. Recent implicit interactive proofs of knowledge can be a starting point. But stronger properties are first expected for improving privacy. They can also be integrated into new ad-hoc broadcast systems, in order to distribute the management among several parties, and eventually remove any trust requirements.

Strong privacy for the Cloud would have a huge societal impact since it would revolutionize the trust model: users would be able to make safe use of outsourced storage, namely for personal, financial and medical data, without having to worry about failures or attacks of the server.

4.2. Hardware Security

Cryptography is only one component of information security, but it is a crucial component. Without cryptography, it would be impossible to establish secure communications between users over insecure networks like the Internet. In particular, public-key cryptography (invented by Diffie and Hellman in 1976) enables to establish secure communications between users who have never met physically before. One can argue that companies like E-Bay or Amazon could not exist without public-key cryptography. Since 30 years the theory of cryptography has developed considerably. However cryptography is not only a theoretical science; namely at some point the cryptographic algorithms must be implemented on physical devices, such as PCs, smart cards or RFIDs. Then problems arise: in general smart cards and RFIDs have limited computing power and leak information through power consumption and electro-magnetic radiations. Similarly a PC can be exposed to various computer viruses which can leak private informations to a remote attacker. Such information leakage can be exploited by an attacker; this is called a **side-channel attack**. It is well known that a cryptographic algorithm which is perfectly secure in theory can be completely insecure in practice if improperly implemented.

In general, countermeasures against side-channel attacks are heuristic and can only make a particular implementation resist particular attacks. Instead of relying on ad-hoc security patches, a better approach consists in working in the framework of **provable security**. The goal is to prove that a cryptosystem does not only resist specific attacks but can resist any possible side-channel attack. As already demonstrated with cryptographic protocols, this approach has the potential to significantly increase the security level of cryptographic products. Recently the cryptography research community has developed new security models to take into account these practical implementation attacks; the most promising such model is called the **leakage-resilient model**.

Therefore, our goal is to define new security models that take into account any possible side-channel attack, and then to design new cryptographic schemes and countermeasures with a proven security guarantee against side-channel attacks.

4.3. Copyright Protection

Similarly to the privacy concern, the digital world makes easy the large-scale diffusion of information. But in some cases, this can be used in violation of some copyrights. Cryptography should help at solving this problem, which is actually two-fold: one can either mark the original document in order to be able to follow the distribution (and possibly trace the traitor who illegally made it public) or one can publish information in an encrypted way, so that authorized people only can access it.

CLASSIC Project-Team

4. Application Domains

4.1. Forecasting of the electricity consumption

Our partner is EDF R&D. The goal is to aggregate in a sequential fashion the forecasts made by some (about 20) base experts in order to predict the electricity consumption at a global level (the one of all French customers) at a half-hourly step. We need to abide by some operational constraints: the predictions need to be made at noon for the next 24 hours (i.e., for the next 48 time rounds).

4.2. Forecasting of the air quality

Our partner is the Inria project-team CLIME (Paris-Rocquencourt). The goal is to aggregate in a sequential fashion the forecasts made by some (about 100) base experts in order to output field prediction of the concentration of some pollutants (typically, the ozone) over Europe. The results were and will be transferred to the public operator INERIS, which uses and will use them in an operational way.

4.3. Forecasting of the production data of oil reservoirs

Our partner is IFP Energies nouvelles. The goal is to aggregate in a sequential fashion the forecasts made by some (about 100) base experts in order to predict some behaviors (gas/oil ratio, cumulative oil extracted, water cut) of the exploitation of some oil wells.

4.4. Forecasting of exchange rates

Our partner is HEC Paris. The goal is to aggregate in a sequential fashion the forecasts made by some (about 5) base macro-economic variables to predict monthly-averaged exchange rates.

4.5. Data mining, massive data sets

Our partner is the start-up Safety Line. The purpose of this application is to investigate statistical learning strategies for mining massive data sets originated from aircraft high-frequency recordings and improve security.

4.6. Computational linguistics

We propose and study new language models that bridge the gap between models oriented towards the statistical analysis of large corpora and grammars oriented towards the description of syntactic features as understood by academic experts. We have conceived a new kind of grammar, based on some cut and paste mechanism and some label aggregation principle, that can be fully learnt from a corpus. We are currently testing this model and studying its mathematical properties and relations with some other new statistical models based on conditional independence assumptions.

4.7. Statistical inference on biological data

The question is about understanding how interactions between neurons can be detected. A mathematical modeling is given by multivariate Hawkes processes. Lasso-type methods can then be used to estimate interaction functions in the nonparametric setting by using fast algorithms, providing inference of the unitary event activity of individual neurons.

CLIME Project-Team

4. Application Domains

4.1. Introduction

The central application domain of the project-team is atmospheric chemistry. We develop and maintain the air quality modeling system Polyphemus, which includes several numerical models (Gaussian models, Lagrangian model, two 3D Eulerian models including Polair3D) and their adjoints, and different high level methods: ensemble forecast, sequential and variational data assimilation algorithms. Advanced data assimilation methods, network design, inverse modeling, ensemble forecast are studied in the context of air chemistry. Note that addressing these high level issues requires controlling the full software chain (models and data assimilation algorithms).

The activity on assimilation of satellite data is mainly carried out for meteorology and oceanography. This is addressed in cooperation with external partners who provide numerical models. Concerning oceanography, the aim is to assess ocean surface circulation, by assimilating fronts and vortices displayed on image acquisitions. Concerning meteorology, the focus is on correcting the model location of structures related to high-impact weather events (cyclones, convective storms, etc.) by assimilating images.

4.2. Air quality

Air quality modeling implies studying the interactions between meteorology and atmospheric chemistry in the various phases of matter, which leads to the development of highly complex models. The different usages of these models comprise operational forecasting, case studies, impact studies, etc., with both societal (e.g., public information on pollution forecast) and economical impacts (e.g., impact studies for dangerous industrial sites). Models lack some appropriate data, for instance better emissions, to perform an accurate forecast and data assimilation techniques are recognized as a major key point for improving forecast's quality.

In this context, Clime is interested in various problems, the following being the crucial ones:

- The development of ensemble forecast methods for estimating the quality of the prediction, in relation with the quality of the model and the observations. This allows sensitivity analysis with respect to the model's parameters so as to identify physical and chemical processes, whose modeling must be improved.
- The development of methodologies for sequential aggregation of ensemble simulations. What ensembles should be generated for that purpose, how spatialized forecasts can be generated with aggregation, how can the different approaches be coupled with data assimilation?
- The definition of second-order data assimilation methods for the design of optimal observation networks. The two main objectives are: management of combinations of sensor types and deployment modes and dynamic management of mobile sensors' trajectories.
- How to estimate the emission rate of an accidental release of a pollutant, using observations and a dispersion model (from the near-field to the continental scale)? How to optimally predict the evolution of a plume? Hence, how to help people in charge of risk evaluation for the population?
- The definition of non-Gaussian approaches for data assimilation.
- The assimilation of satellite measurements of troposphere chemistry.

The activities of Clime in air quality are supported by the development of the Polyphemus air quality modeling system. This system has a modular design, which makes it easier to manage high level applications such as inverse modeling, data assimilation and ensemble forecast.

4.3. Oceanography

The capacity of performing a high quality forecast of the state of the ocean, from the regional to the global scales, is of major interest. Such a forecast can only be obtained by systematically coupling numerical models and observations (in situ and satellite data). In this context, being able to assimilate image structures becomes a key point. Examples of such image structures are:

- apparent motion that represents surface velocity;
- trajectories, obtained either from tracking of features or from integration of the velocity field;
- spatial objects, such as fronts, eddies or filaments.

Image models of these structures are developed and take into account the underlying physical processes. Image data are assimilated in these image models to derive pseudo-observations of state variables, which are further assimilated in numerical ocean forecast models.

4.4. Meteorology

Meteorological forecasting constitutes a major applicative challenge for image assimilation. Although satellite data are operationally assimilated within models, this is mainly done on an independent pixel basis: the observed radiance is linked to the state variables via a radiative transfer model, that plays the role of an observation operator. Indeed, because of their limited spatial and temporal resolutions, numerical weather forecast models fail to exploit image structures, such as precursors of high impact weather:

- cyclogenesis related to the intrusion of dry stratospheric air in the troposphere (a precursor of cyclones),
- convective systems (supercells) leading to heavy winter time storms,
- low-level temperature inversion leading to fog and ice formation, etc.

To date, there is no available method for assimilating such data, which are characterized by a strong coherence in space and time. Meteorologists have developed qualitative Conceptual Models (CMs), for describing the high impact weathers and their signature on images, and tools to detect CMs on image data. The result of this detection is used for correcting the numerical models, for instance by modifying the initialization. The aim is therefore to develop a methodological framework allowing to assimilate the detected CMs within numerical forecast models. This is a challenging issue given the considerable impact of the related meteorological events.

CONTRAINTES Project-Team

4. Application Domains

4.1. Combinatorial optimization

The number and economic impact of combinatorial optimization problems found in the industrial world are constantly increasing. They cover:

- resource allocation;
- placement, bin packing;
- scheduling;
- planning;
- transport;
- etc.

The last fifty years have brought many improvements in Operations Research resolution techniques. In this context, Constraint Programming can be seen as providing, on the one hand, constraint propagation algorithms that can be applied to various numerical or symbolic constraints, and on the other hand, declarative languages to model real-life problems and express complex resolution strategies. The latter point is crucial for designing new algorithms that cannot be defined without a sufficiently high-level language to express them. It allowed for better results than traditional methods, for instance in scheduling, and is promised to an even better future when thinking about the cooperation of global resolution, local consistency techniques and search methods.

The European FP6 Strep project Net-WMS that we have coordinated, has shown the benefit of combining discrete geometry constraints with rules to express physical, common sense and packing business constraints to solve packing problems in the context of warehouse management systems for the automotive industry. In this context, we have developed a rule-based modeling language, called Rules2CP, to express requirements in a declarative and flexible manner, and compile them to efficient constraint programs using reified constraints and a global constraint dedicated to geometrical placement problems in high dimension.

4.2. Computational Systems Biology

In partnership with biologists, we develop and experiment our modeling methods in five main leading applications:

- **Cancer chronotherapy optimization.** This research initiated in 2004 in partnership with Jean Clairambault, EPI BANG, and Francis Lévi INSERM, Hopital Paul Brousse, Villejuif, aims at understanding fundamental mechanisms involved in cancer and chronotherapies through mathematical modeling. Following the EU STREP project TEMPO (2006-2009) on “temporal genomics for patient tailored chronotherapeutics”, coordinated by Francis Lévi, and in the framework of the Era-Net SysBio C5Sys project (2010-2013) coordinated by Francis Lévi and David Rand, University of Warwick, UK, we develop coupled models of the cell cycle, the circadian clock, the DNA repair system, irinotecan metabolism and drug injection optimization, focussing on the interactions between the cell cycle and the circadian clock in mammalian cells.
- **Mammalian cell cycle regulation.** This theme that is closely related to the previous one has lead to a formal collaboration in the framework of the ANR Syscomm project CALAMAR, started in 2009 on the “Compositional modeling and Analysis of LArge MoleculAr Regulatory networks”. In partnership with Claudine Chaouiya, TAGC INSERM, Marseille, and Laurence Calzone, Institut Curie, Paris, this project aims at applying our computational techniques – both qualitative and quantitative – to the analysis of the large scale RB/E2F network, in order to elucidate various features of the human cell proliferation, especially in the case of healthy and bladder-tumor cells of different aggressiveness.

- **Real-time control of gene expression in yeast.** This research lead in the team by Grégory Batt investigates the possibilities to control gene expression in living cells. In collaboration with Pascal Hersen and Samuel Bottani, biophysicists at the Matière and Systèmes Complexes lab, CNRS/Paris Diderot University, we develop a microfluidic platform and control software for the real-time control of gene expression in yeast. In a larger initiative, we consider a similar problem but in mammalian cells, where the stochasticity of gene expression makes the control problem particularly challenging. The Iceberg Investissement d'Avenir project, coordinated by Grégory Batt, involves the MSC, BM2A, LIFL and PPS labs, and the Jacques Monod Institut. Similarly, the Contraintes research group is also involved in the Inria/INSERM large-scale initiative action COLAGE coordinated by Huges Berry, EPI COMBINING, with François Taddei, Ariel Lindner, INSERM Paris Necker, Hidde de Jong, Delphine Ropers, EPI IBIS, Jean-Luc Gouzé, and Madalena Chaves, EPI COMORE. In this project, we investigate the possibilities to control and reprogram growth and aging in bacteria *E. coli* using synthetic biology approaches.
- **Artificial tissue homeostasis in mammalian cells.** Artificial tissue design is a particularly challenging problem in synthetic biology since the system behavior results from the interplay between intra- and intercellular dynamics. In the framework of the Syne2arti ANR project, coordinated by Grégory Batt, and involving Dirk Draso, EPI BANG, Oded Maler, CNRS Verimag, and Ron Weiss, MIT, USA, we design and genetically-engineer mammalian cells to obtain a tissue having a desired cell density. The long-term correct functioning of the system relies several key aspects, including individual cell decisions, collective, spatial aspects, and cell-to-cell variability.
- **TGFβ signaling** In the framework of the BioTempo ANR project, we recently started to apply the different algorithms available in the BIOCHAM platform to the modeling of the TGFβ signaling network in collaboration with the SeRAIC lab (Rennes, France). The main challenge is to compare and understand crosstalks between the SMAD-dependent fast pathway and the MAPK-dependent slower pathway that is often related to cancer. Both the static network analyzers and the parameter learning methods of BIOCHAM are put to good use in this work.

CRYPT Team

4. Application Domains

4.1. Security Estimates for Cryptography

An important application of cryptanalysis is to evaluate the concrete security of a given cryptosystem, so that key sizes and parameters are chosen appropriately. In some sense, cryptanalysis is the crash test of cryptography. When one uses cryptography, the first thing that one does is to select parameters and key sizes: in the real world, several well-known cryptographic failures happened due to inappropriate key sizes. Cryptanalysis analyzes the best attacks known: it assesses their cost (depending on the platform) and their performances (such as success probability). Sometimes the exact cost of an attack cannot be evaluated accurately nor rigorously, but fortunately, it is often possible to give an order of magnitude, which allows to select key sizes with a reasonable security margin.

On the other hand, it must be stressed that cryptanalysis depends on the state of the art: today's best attack may be completely different from tomorrow's best attack. The case of MD5 is a good reminder of this well-known fact.

4.2. Algorithmic Number Theory

Algorithms developed for cryptanalysis have sometimes applications outside cryptanalysis, especially in algorithmic number theory. This has happened for lattices and elliptic curves, and is not surprising, considering that some of the problems studied by cryptanalysis are very basic (like integer factoring), and therefore ubiquitous. Cryptanalysis motivates the search of truly-efficient algorithms, and experiments are common in public-key cryptanalysis, which allows to really verify improvements.

DEDUCTEAM Exploratory Action

4. Application Domains

4.1. Safety of Aerospace systems

In parallel with this effort in logic and in the development of proof checkers and automated theorem proving systems, we always have been interested in using such tools. One of our favorite application domain is the safety of aerospace systems. Together with César Muñoz' team in Nasa-Langley, we have proved the correctness of several geometric algorithms used in air traffic control.

This has led us sometimes to develop such algorithms ourselves, and sometimes to develop tools for automating these proofs.

4.2. Tools for proofs in B

Set theory appears to be an appropriate theory for automated theorem provers based on Deduction modulo, in particular the several extensions of *Zenon* (*Super Zenon* and *Zenon Modulo*). Modeling techniques using set theory are therefore good candidates to assess these tools. This is what we have done with the *B* method whose formalism relies on set theory. A collaboration with *Siemens* has been developed to automatically verify the *B* proof rules of *Atelier B* [10]. From this work presented in the Doctoral dissertation of Mélanie Jacquel, the *Super Zenon* tool [5] has been designed in order to be able to reason modulo the *B* set theory. As a sequel of this work, we contribute to the *BWare* project whose aim is to provide a mechanized framework to support the automated verification of *B* proof obligations coming from the development of industrial applications. In this context, we have recently designed *Zenon Modulo* [22], [23] (Pierre Halmagrand's PhD thesis, which has started on October 2013) to deal with the *B* set theory. In this work, the idea is to manually transform the *B* set theory into a theory modulo and provide it to *Zenon Modulo* in order to verify the proof obligations of the *BWare* project.

DYOGENE Project-Team

4. Application Domains

4.1. Embedded Networks

Critical real-time embedded systems (cars, aircrafts, spacecrafts) are nowadays made up of multiple computers communicating with each other. The real-time constraints typically associated with operating systems now extend to the networks of communication between sensors/actuators and computers, and between the computers themselves. Once a media is shared, the time between sending and receiving a message depends not only on technological constraints, but also, and mainly from the interactions between the different streams of data sharing the media. It is therefore necessary to have techniques to guarantee maximum network delays, in addition to local scheduling constraints, to ensure a correct global real-time behaviour to distributed applications/functions.

Moreover, pessimistic estimate may lead to an overdimensioning of the network, which involves extra weight and power consumption. In addition, these techniques must be scalable. In a modern aircraft, thousands of data streams share the network backbone. Therefore algorithm complexity should be at most polynomial.

4.2. Routing protocols

Routing protocols enables to maintain paths for transmitting messages over a network. Those protocols, such as OSPF, are based on the transmission of periodic messages between neighbors. Nowadays, faulty behaviors result in the raising of alarms, but are mostly detected when a breakdown or a major misbehavior occurs. Indeed, alarms are so numerous that thay cannot be analyzed efficiently. We aim at developing methods to detect misbehaviours of a router befor a major fault accurs, and techniques to study the influence of the protocol parameters on the bahavior of the network.

4.3. Wireless Networks

Wireless networks can be efficiently modelled as dynamic stochastic geometric networks. Their analysis requires taking into account, in addition to their geometric structure, the specific nature of radio channels and their statistical properties which are often unknown a priori, as well as the interaction through interference of the various individual point-to-point links.

4.4. Peer-to-Peer Systems

The amount of multimedia traffic accessed via the Internet, already of the order of exabytes (10^{18} bytes) per month, is expected to grow steadily in the coming years. A peer-to-peer (P2P) architecture, where peers contribute resources to support service of such traffic, holds the promise to support its growth more cheaply than by scaling up the size of data centers. More precisely, a large scale P2P system based on resources of individual users can absorb part of the load that would otherwise need to be served by data centers. In video-on-demand applications, the critical resources at the peers are storage space and uplink bandwidth. Our objective is to ensure that the largest fraction of traffic is supported by the P2P system.

4.5. Social and economic networks

Networks are ubiquitous with the presence of different kinds of social, economic and information networks around us. The Internet is one of the most prominent examples of a geometric network. We also examine geometric networks from the perspective of sociologist and economist [70]. Network analysis is also attracting foundational research by computer scientists [63]. Diffusion of information, social influence, trust, communication and cooperation between agents are heavily researched topics in e-commerce and multi-agent systems. Our probabilistic techniques are very appropriate in this case and have been largely neglected so far. While the first works on geometric networks emanated from theoretical physicists, they stay more focused on static properties of such networks and do not consider game theoretical or statistical learning (like community detection) aspects of such networks. This leaves open a range of new problems to which we will contribute.

FORMES Team

4. Application Domains

4.1. Proof of Programs

In many life critical application such as nuclear power or transportation, formal proofs of programs are required, and theorem provers provide an essential tool in that area.

4.2. Simulation

Simulation is relevant to most areas where complex embedded systems are used, not only to the semiconductor industry for System-on-Chip modeling, but also to any application where a complex hardware platform must be assembled to run the application software. It has applications for example in industry automation, digital TV, telecommunications and transportation.

4.3. Certified Compilation for Embedded systems

Many frameworks have been designed in order to make the design and the development of embedded systems more rigourous and secure on the basis of some formal model. All these frameworks implicitly assume the *reliability of the translation* to executable code, in order to guarantee the verified properties in the design level are preserved in the implementation. In other words, they rely on a claim saying that the compilers from high level model description to the implementation will not introduce undesired behaviors or errors in silence. The only safe way to satisfy such a claim is to certify correctness of the compilers, that is, to prove that the code they produce has exactly the semantics of the source code or model.

4.4. Distributed Systems

Many embedded systems run in a distributed environment. Distributed systems raise extremely challenging issues, both for the design and the implementation, because decisions can be made only from a local knowledge, which is imperfect due to communication time and unreliability of transmissions.

4.5. Security

The convergence between embedded technologies and the Internet offers many opportunities to malicious people for breaking the privacy of consumers or of organisations. Using cryptography is not enough for ensuring the protection of data, because of possible flaws in protocols and interfaces, providing opportunities for many well-known attacks. This area is therefore an important target of formal methods.

GALLIUM Project-Team

4. Application Domains

4.1. High-assurance software

A large part of our work on programming languages and tools focuses on improving the reliability of software. Functional programming, program proof, and static type-checking contribute significantly to this goal.

Because of its proximity with mathematical specifications, pure functional programming is well suited to program proof. Moreover, functional programming languages such as Caml are eminently suitable to develop the code generators and verification tools that participate in the construction and qualification of high-assurance software. Examples include Esterel Technologies's KCG 6 code generator, the Astrée static analyzer, the Caduceus/Jessie program prover, and the Frama-C platform. Our own work on compiler verification combines these two aspects of functional programming: writing a compiler in a pure functional language and mechanically proving its correctness.

Static typing detects programming errors early, prevents a number of common sources of program crashes (null references, out-of bound array accesses, etc), and helps tremendously to enforce the integrity of data structures. Judicious uses of generalized abstract data types (GADTs), phantom types, type abstraction and other encapsulation mechanisms also allow static type checking to enforce program invariants.

4.2. Software security

Static typing is also highly effective at preventing a number of common security attacks, such as buffer overflows, stack smashing, and executing network data as if it were code. Applications developed in a language such as Caml are therefore inherently more secure than those developed in unsafe languages such as C.

The methods used in designing type systems and establishing their soundness can also deliver static analyses that automatically verify some security policies. Two examples from our past work include Java bytecode verification [45] and enforcement of data confidentiality through type-based inference of information flows and noninterference properties [49].

4.3. Processing of complex structured data

Like most functional languages, Caml is very well suited to expressing processing and transformations of complex, structured data. It provides concise, high-level declarations for data structures; a very expressive pattern-matching mechanism to destructure data; and compile-time exhaustiveness tests. Languages such as CDuce and OCamlDuce extend these benefits to the handling of semi-structured XML data [39]. Therefore, Caml is an excellent match for applications involving significant amounts of symbolic processing: compilers, program analyzers and theorem provers, but also (and less obviously) distributed collaborative applications, advanced Web applications, financial modeling tools, etc.

4.4. Rapid development

Static typing is often criticized as being verbose (due to the additional type declarations required) and inflexible (due to, for instance, class hierarchies that must be fixed in advance). Its combination with type inference, as in the Caml language, substantially diminishes the importance of these problems: type inference allows programs to be initially written with few or no type declarations; moreover, the OCaml approach to object-oriented programming completely separates the class inheritance hierarchy from the type compatibility relation. Therefore, the Caml language is highly suitable for fast prototyping and the gradual evolution of software prototypes into final applications, as advocated by the popular "extreme programming" methodology.

4.5. Teaching programming

Our work on the Caml language has an impact on the teaching of programming. Caml Light is one of the programming languages selected by the French Ministry of Education for teaching Computer Science in *classes préparatoires scientifiques*. OCaml is also widely used for teaching advanced programming in engineering schools, colleges and universities in France, the USA, and Japan.

GAMMA3 Project-Team (section vide)

GANG Project-Team

4. Application Domains

4.1. Application Domains

Application domains include evaluating Internet performances, the design of new peer-to-peer applications, enabling large scale ad hoc networks and mapping the web.

- The application of measuring and modeling Internet metrics such as latencies and bandwidth is to provide tools for optimizing Internet applications. This concerns especially large scale applications such as web site mirroring and peer-to-peer applications.
- Peer-to-peer protocols are based on a all equal paradigm that allows to design highly reliable and scalable applications. Besides the file sharing application, peer-to-peer solutions could take over in web content dissemination resistant to high demand bursts or in mobility management. Envisioned peer-to-peer applications include video on demand, streaming, exchange of classified ads,...
- Wifi networks have entered our every day life. However, enabling them at large scale is still a challenge. Algorithmic breakthrough in large ad hoc networks would allow to use them in fast and economic deployment of new radio communication systems.
- The main application of the web graph structure consists in ranking pages. Enabling site level indexing and ranking is a possible application of such studies.

HIPERCOM2 Team

4. Application Domains

4.1. Introduction

The HIPERCOM2 team addresses the following application domains:

- military, emergency or rescue applications,
- industrial applications,
- vehicular networks,
- smart cities,
- Internet of Things.

These application domains use the four types of wireless networks:

- wireless mesh and mobile ad hoc networks,
- wireless sensor networks,
- vehicular networks,
- cognitive radio networks.

4.2. Wireless mesh and mobile ad hoc networks

A mobile ad hoc network is a network made of a collection of mobile nodes that gather spontaneously and communicate without requiring a pre-existing infrastructure. Of course a mobile ad hoc network use a wireless communication medium. They can be applied in various contexts:

- military;
- rescue and emergency;
- high speed access to internet.

The military context is historically the first application of mobile ad hoc networks.

The rescue context is halfway between military and civilian applications. In emergency applications, heterogeneous wireless networks have to cooperate in order to save human lifes or bring the situation back to normal as soon as possible. Wireless networks that can be quickly deployed are very useful to assess damages and take the first decisions appropriate to the disaster of natural or human origin. The primary goal is to maintain connectivity with the humans or the robots (in case of hostile environment) in charge of network deployment. This deployment should ensure the coverage of an interest area or of only some interest points. The wireless network has to cope with pedestrian mobility and robots/vehicles mobility. The environment, initially unknown, is progressively discovered and usually has many obstacles. These obstacles should be avoided. The nodes of the wireless network are usually battery-equipped. Since they are dropped by a robot or a human, their weight is very limited. The protocols supported by these nodes should be energy efficient to increase network lifetime. Furthermore, in case of aggressive environment, sensor nodes should be replaced before failing. Hence, in such conditions, it is required to predict the failure time of nodes to favor a predictive maintenance.

Mobile ad hoc network provide an enhanced coverage for high speed wireless access to the internet. The now very popular WLAN standard, WiFi, provides much larger capacity than mobile operator networks. Using a mobile ad hoc network around hot spots will offer high speed access to much larger community, including cars, busses, trains and pedestrians.

4.3. Vehicular Networks and Smart Cities

Vehicular ad hoc networks (VANET) are based on short- to medium-range transmission systems that support both vehicle-to-vehicle and vehicle-to-roadside communications. Vehicular networks will enable vehicular safety applications (safety warnings) as well as non-safety applications (real-time traffic information, routing support, mobile entertainment, and many others). We are interested in developing an efficient routing protocol that takes advantage of the fixed network infrastructure deployed along the roads. We are also studying MAC layer issues in order to provide more priority for security messages which have stringent delivery constraints.

Smart cities share with the military tactical networks the constraint on pedestrian and vehicular mobility. Furthermore, the coexistence of many networks operating in the same radio spectrum may cause interferences that should be avoided. Cognitive radio takes advantage of the channels temporarily left available by the primary users to assign them to secondary users. Such an opportunistic behavior can also be applied in wireless sensor networks deployed in the cities. Smart cities raise the problem of transmitting, gathering, processing and storing big data. Another issue is to provide the right information at the right place: where it is needed.

4.4. Wireless sensor networks in industrial applications and Internet of Things

Concerning wireless sensor networks, WSNs, we tackle the three following issues:

- Energy efficiency is a key property in wireless sensor networks. Various techniques contribute to save energy of battery-equipped sensor nodes. To name a few, they are: energy efficient routing protocols, node activity scheduling, adjustment of transmision power, reduction of protocols overhead, reduction of data generated and transmitted. In the OCARI network, an industrial wireless sensor network, we have designed and implemented an energy efficient routing protocol and a node activity scheduling algorithm allowing router nodes to sleep. We have applied a cross-layering approach allowing the optimization of MAC and network protocols taking into account the application requirements and the environment in which the network operates. We have observed the great benefit obtained with node activity scheduling. In networks with low activity, opportunistic strategies are used to address low duty cycles.
- Large scale WSNs constitute another challenge. Large autonomous wireless sensors in the internet of the things need very well tuned algorithms. Self-organization is considered as a key element in tomorrow's Internet architecture. A major challenge concerning the integration of self-organized networks in the Internet is the accomplishment of light weight network protocols in large ad hoc environments.
- Multichannel WSNs provide an opportunity:
  - on the one hand, to increase the parallelism between transmissions. Hence, it reduces the data gathering delays and improves the time consitency of gathered data.
  - on the other hand, to increase the robustness against interferences and perturbations possibly caused by the coexistence of other wireless networks.

4.5. Cognitive Radio Networks

Usually in cognitive radio, the secondary users are in charge of monitoring the channel to determine whether or not the primary users are active in the area. If they are not, the secondary users are allowed to use the spectrum left unused by the primary users. We are interested in two issues:

- Design and modeling of a new access scheme based on a generalized Carrier Sense Multiple Access scheme using active signaling. This scheme allows the primary users to capture the bandwidth even if the secondary users are transmitting in the area.
- Design of a time slot and channel assignment to minimize the data gathering performed by secondary users. This assignment should work with different detection schemes of primary user presence.

IMARA Project-Team

4. Application Domains

4.1. Introduction

While the preceding section focused on methodology, in connection with automated guided vehicles, it should be stressed that the evolution of the problems which we deal with, remains often guided by the technological developments. We enumerate three fields of application, whose relative importance varies with time and which have strong mutual dependencies: driving assistance, cars available in self-service mode and fully automated vehicles (cybercars).

4.2. Driving assistance

Several techniques will soon help drivers. One of the first immediate goal is to improve security by alerting the driver when some potentially dangerous or dangerous situations arise, i.e. collision warning systems or lane tracking could help a bus driver and surrounding vehicle drivers to more efficiently operate their vehicles. Human factors issues could be addressed to control the driver workload based on additional information processing requirements.

Another issue is to optimize individual journeys. This means developing software for calculating optimal (for the user or for the community) paths. Nowadays, path planning software is based on a static view of the traffic: efforts have to be done to take the dynamic component in account.

4.3. New transportation systems

The problems related to the abusive use of the individual car in large cities led the populations and the political leaders to support the development of public transport. A demand exists for a transport of people and goods which associates quality of service, environmental protection and access to the greatest number. Thus the tram and the light subways of VAL type recently introduced into several cities in France conquered the populations, in spite of high financial costs.

However, these means of mass transportation are only possible on lines on which there is a keen demand. As soon as one moves away from these “lines of desire” or when one deviates from the rush hours, these modes become expensive and offer can thus only be limited in space and time.

To give a more flexible offer, it is necessary to plan more individual modes which approach the car as we know it. However, if one wants to enjoy the benefits of the individual car without suffering from their disadvantages, it is necessary to try to match several criteria: availability anywhere and anytime to all, lower air and soils pollution as well as sound levels, reduced ground space occupation, security, low cost.

Electric or gas vehicles available in self-service, as in the Praxitèle system, bring a first response to these criteria. To be able to still better meet the needs, it is however necessary to re-examine the design of the vehicles on the following points:

- ease empty car moves to better distribute them;
- better use of information systems inboard and on ground;
- better integrate this system in the global transportation system.

These systems are now operating (i.e. in La Rochelle). The challenge is to bring them to an industrial phase by transferring technologies to these still experimental projects.

4.4. Automated vehicles

The long term effort of the project is to put automatically guided vehicles (cybercars) on the road. It seems too early to mix cybercars and traditional vehicles, but data processing and automation now make it possible to consider in the relatively short term the development of such vehicles and the adapted infrastructures. IMARA aims at using these technologies on experimental platforms (vehicles and infrastructures) to accelerate the technology transfer and to innovate in this field.

Other application can be precision docking systems that will allow buses to be automatically maneuvered into a loading zone or maintenance area, allowing easier access for passengers, or more efficient maintenance operations. Transit operating costs will also be reduced through decreased maintenance costs and less damage to the braking and steering systems.

Regarding technical topics, several aspects of Cybercars have been developed at IMARA this year. First, we have stabilized a generic Cycab architecture involving Inria Syndex tool and CAN communications. The critical part of the vehicle is using a real-time Syndex application controlling the actuators via two Motorola's MPC555. Today, we have decided to migrate to the new dsPIC architecture for more efficiency and ease of use.

This application has a second feature, it can receive commands from an external source (Asynchronously this time) on a second CAN bus. This external source can be a PC or a dedicated CPU, we call it high level. To work on the high level, in the past years we have been developing a R&D framework called (Taxi) which used to take control of the vehicle (Cycab and Yamaha) and process data such as gyro, GPS, cameras, wireless communications and so on. Today, in order to rely on a professional and maintained solution, we have chosen to migrate to the RTMaps SDK development platform. Today, all our developments and demonstrations are using this efficient prototyping platform. Thanks to RTMaps we have been able to do all the demonstrations on our cybercars: cycabs, Yamaha AGV and new Cybus platforms. These demonstrations include: reliable SLAMMOT algorithm using 2 to 4 laser sensors simultaneously, automatic line/road following techniques, PDA remote control, multi sensors data fusion, collaborative perception via ad-hoc network.

The second main topic is inter-vehicle communications using ad-hoc networks. We have worked with the HIPERCOM team for setting and tuning OLSR, a dynamic routing protocol for vehicles communications (see Section 3.2). Our goal is to develop a vehicle dedicated communication software suite, running on a specialized hardware. It can be linked also with the Taxi Framework for getting data such GPS information's to help the routing algorithm.

MATHRISK Project-Team

4. Application Domains

4.1. Application Domains

Risk management, Quantitative Finance, Computational Finance, Market Microstructure, Systemic Risk, Portfolio optimization, Risk modeling.

MICMAC Project-Team

4. Application Domains

4.1. Electronic structure of large systems

As the size of the systems one wants to study increases, more efficient numerical techniques need to be resorted to. In computational chemistry, the typical scaling law for the complexity of computations with respect to the size of the system under study is N^3, N being for instance the number of electrons. The Holy Grail in this respect is to reach a linear scaling, so as to make possible simulations of systems of practical interest in biology or material science. Efforts in this direction must address a large variety of questions such as

- how can one improve the nonlinear iterations that are the basis of any *ab initio* models for computational chemistry?
- how can one more efficiently solve the inner loop which most often consists in the solution procedure for the linear problem (with frozen nonlinearity)?
- how can one design a sufficiently small variational space, whose dimension is kept limited while the size of the system increases?

An alternative strategy to reduce the complexity of *ab initio* computations is to try to couple different models at different scales. Such a mixed strategy can be either a sequential one or a parallel one, in the sense that

- in the former, the results of the model at the lower scale are simply used to evaluate some parameters that are inserted in the model for the larger scale: one example is the parameterized classical molecular dynamics, which makes use of force fields that are fitted to calculations at the quantum level;
- while in the latter, the model at the lower scale is concurrently coupled to the model at the larger scale: an instance of such a strategy is the so called QM/MM coupling (standing for Quantum Mechanics/Molecular Mechanics coupling) where some part of the system (typically the reactive site of a protein) is modeled with quantum models, that therefore accounts for the change in the electronic structure and for the modification of chemical bonds, while the rest of the system (typically the inert part of a protein) is coarse grained and more crudely modeled by classical mechanics.

The coupling of different scales can even go up to the macroscopic scale, with methods that couple a microscopic description of matter, or at least a mesoscopic one, with the equations of continuum mechanics at the macroscopic level.

4.2. Computational Statistical Mechanics

The orders of magnitude used in the microscopic description of matter are far from the orders of magnitude of the macroscopic quantities we are used to: The number of particles under consideration in a macroscopic sample of material is of the order of the Avogadro number $\mathcal{N}_A \sim 10^{23}$, the typical distances are expressed in Å (10^{-10} m), the energies are of the order of $k_\mathrm{B}T \simeq 4 \times 10^{-21}$ J at room temperature, and the typical times are of the order of 10^{-15} s when the proton mass is the reference mass.

To give some insight into such a large number of particles contained in a macroscopic sample, it is helpful to compute the number of moles of water on earth. Recall that one mole of water corresponds to 18 mL, so that a standard glass of water contains roughly 10 moles, and a typical bathtub contains 10^5 mol. On the other hand, there are approximately 1.3×10^{18} m^3 of water in the oceans, *i.e.*7.2×10^{22} mol, a number comparable to the Avogadro number. This means that inferring the macroscopic behavior of physical systems described at the microscopic level by the dynamics of several millions of particles only is like inferring the ocean's dynamics from hydrodynamics in a bathtub...

For practical numerical computations of matter at the microscopic level, following the dynamics of every atom would require simulating $\mathcal{N}_A$ atoms and performing $\mathrm{O}(10^{15})$ time integration steps, which is of course impossible! These numbers should be compared with the current orders of magnitude of the problems that can be tackled with classical molecular simulation, where several millions of atoms only can be followed over time scales of the order of 0.1 μs.

Describing the macroscopic behavior of matter knowing its microscopic description therefore seems out of reach. Statistical physics allows us to bridge the gap between microscopic and macroscopic descriptions of matter, at least on a conceptual level. The question is whether the estimated quantities for a system of N particles correctly approximate the macroscopic property, formally obtained in the thermodynamic limit $N \to +\infty$ (the density being kept fixed). In some cases, in particular for simple homogeneous systems, the macroscopic behavior is well approximated from small-scale simulations. However, the convergence of the estimated quantities as a function of the number of particles involved in the simulation should be checked in all cases.

Despite its intrinsic limitations on spatial and timescales, molecular simulation has been used and developed over the past 50 years, and its number of users keeps increasing. As we understand it, it has two major aims nowadays.

First, it can be used as a *numerical microscope*, which allows us to perform “computer” experiments. This was the initial motivation for simulations at the microscopic level: physical theories were tested on computers. This use of molecular simulation is particularly clear in its historic development, which was triggered and sustained by the physics of simple liquids. Indeed, there was no good analytical theory for these systems, and the observation of computer trajectories was very helpful to guide the physicists’ intuition about what was happening in the system, for instance the mechanisms leading to molecular diffusion. In particular, the pioneering works on Monte-Carlo methods by Metropolis et al, and the first molecular dynamics simulation of Alder and Wainwright were performed because of such motivations. Today, understanding the behavior of matter at the microscopic level can still be difficult from an experimental viewpoint (because of the high resolution required, both in time and in space), or because we simply do not know what to look for! Numerical simulations are then a valuable tool to test some ideas or obtain some data to process and analyze in order to help assessing experimental setups. This is particularly true for current nanoscale systems.

Another major aim of molecular simulation, maybe even more important than the previous one, is to compute macroscopic quantities or thermodynamic properties, typically through averages of some functionals of the system. In this case, molecular simulation is a way to obtain *quantitative* information on a system, instead of resorting to approximate theories, constructed for simplified models, and giving only qualitative answers. Sometimes, these properties are accessible through experiments, but in some cases only numerical computations are possible since experiments may be unfeasible or too costly (for instance, when high pressure or large temperature regimes are considered, or when studying materials not yet synthesized). More generally, molecular simulation is a tool to explore the links between the microscopic and macroscopic properties of a material, allowing one to address modelling questions such as “Which microscopic ingredients are necessary (and which are not) to observe a given macroscopic behavior?”

4.3. Homogenization and related problems

Over the years, the project-team has developed an increasing expertise on how to couple models written at the atomistic scale, with more macroscopic models, and, more generally, an expertise in multiscale modelling for materials science.

The following observation motivates the idea of coupling atomistic and continuum description of materials. In many situations of interest (crack propagation, presence of defects in the atomistic lattice, ...), using a model based on continuum mechanics is difficult. Indeed, such a model is based on a macroscopic constitutive law, the derivation of which requires a deep qualitative and quantitative understanding of the physical and mechanical properties of the solid under consideration. For many solids, reaching such an understanding is a challenge, as loads they are submitted to become larger and more diverse, and as experimental observations helping designing such models are not always possible (think of materials used in the nuclear industry). Using an atomistic model in the whole domain is not possible either, due to its prohibitive computational cost. Recall indeed that a macroscopic sample of matter contains a number of atoms on the order of 10^{23}. However, it turns out that, in many situations of interest, the deformation that we are after is not smooth in *only a small part* of the solid. So, a natural idea is to try to take advantage of both models, the continuum mechanics one and the atomistic one, and to couple them, in a domain decomposition spirit. In most of the domain, the deformation is expected to be smooth, and reliable continuum mechanics models are then available. In the rest of the domain, the expected deformation is singular, one needs an atomistic model to describe it properly, the cost of which remains however limited as this region is small.

From a mathematical viewpoint, the question is to couple a discrete model with a model described by PDEs. This raises many questions, both from the theoretical and numerical viewpoints:

- first, one needs to derive, from an atomistic model, continuum mechanics models, under some regularity assumptions that encode the fact that the situation is smooth enough for such a macroscopic model to be a good description of the materials;
- second, couple these two models, e.g. in a domain decomposition spirit, with the specificity that models in both domains are written in a different language, that there is no natural way to write boundary conditions coupling these two models, and that one would like the decomposition to be self-adaptive.

More generally, the presence of numerous length-scales in material science problems represents a challenge for numerical simulation, especially when some *randomness* is assumed on the materials. It can take various forms, and includes defects in crystals, thermal fluctuations, and impurities or heterogeneities in continuous media. Standard methods available in the literature to handle such problems often lead to very costly computations. Our goal is to develop numerical methods that are more affordable. Because we cannot embrace all difficulties at once, we focus on a simple case, where the fine scale and the coarse-scale models can be written similarly, in the form of a simple elliptic partial differential equation in divergence form. The fine scale model includes heterogeneities at a small scale, a situation which is formalized by the fact that the coefficients in the fine scale model vary on a small length scale. After homogenization, this model yields an effective, macroscopic model, which includes no small scale. In many cases, a sound theoretical groundwork exists for such homogenization results. We consider mostly the setting of stochastic homogenization of linear, scalar, second order elliptic PDEs, where analytical formulas for the effective properties are known. The difficulty stems from the fact that they generally lead to prohibitively costly computations. For such a case, simple from the theoretical viewpoint, our aim is to focus on different practical computational approaches to speed-up the computations. One possibility, among others, is to look for specific random materials, relevant from the practical viewpoint, and for which a dedicated approach can be proposed, that is less expensive than the general approach.

MOKAPLAN Exploratory Action

4. Application Domains

4.1. Continuous models in Economy

- As already mentioned the CFD formulation is a limit case of simple variational Mean-Field Games (MFG) [55]. MFG is a new branch of game theory recently developed by J-M. Lasry and P-L. Lions. MFG models aim at describing the limiting behavior of stochastic differential games when the number of players tends to infinity. They are specifically designed to model economic problems where a large number of similar interacting agents try to maximize/minimize a utility/cost function which takes into account global but partial information on the game. The players in these models are individually insignificant but they collectively have a significant impact on the cost of the other players. Dynamic MFG models often lead to a system of PDEs which consists of a backward Hamilton-Jacobi Bellman equation for a value function coupled with a forward Fokker-Planck equation describing the space-time evolution of the density of agents.
- In microeconomics, the *principal-agent problem* [74] with adverse selection plays a distinguished role in the literature on asymmetric information and contract theory (with important contributions from several Nobel prizes such as Mirrlees, Myerson or Spence) and it has many important applications in optimal taxation, insurance, nonlinear pricing. The problem can be reduced to the maximization of an integral functional subject to a convexity constraint This is an unusual calculus of variations problem and the optimal price can only be computed numerically. Recently, following a reformulation of Carlier [11], convexity/well-posedness results of McCann, Figalli and Kim [42], connected to optimal transport theory, showed that there is some hope to numerically solve the problem for general utility functions.
- In [8] a class of games are considered with a continuum of players for which Cournot-Nash equilibria can be obtained by the minimisation of some cost, related to optimal transport. This cost is not convex in the usual sense in general but it turns out to have hidden strict convexity properties in many relevant cases. This enables us to obtain new uniqueness results and a characterisation of equilibria in terms of some partial differential equations, a simple numerical scheme in dimension one as well as an analysis of the inefficiency of equilibria. The mathematical problem has the structure of one step of the JKO gradient flow method.
- Many relevant markets are markets of indivisible goods characterized by a certain quality: houses, jobs, marriages... On the theoretical side, recent papers by Ekeland, McCann, Chiappori [34] showed that finding equilibria in such markets is equivalent to solving a certain optimal transport problem (where the cost function depends on the sellers and buyers preferences). On the empirical side, this allows for trying to recover information on the preferences from observed matching; this is an inverse problem as in a recent work of Galichon and Salanié [47] [48] Interestingly, these problems naturally lead to numerically challenging variants of the Monge-Kantorovich problem: the multi-marginal OT problem and the entropic approximation of the Monge-Kantorovich problem (which is actually due to Schrödinger in the early 30's).

4.2. Finance

The Skorohod embedding problem (SEP) consists in finding a martingale interpolation between two probability measures. When a particular stochastic ordering between the two measures is given, Galichon et al [46] have shown that a very natural variational formulation could be given to a class of problems that includes the SEP. This formulation is related to the CFD formulation of the OT problem [2] and has applications to *model-free bounds of derivative prices in Finance*. It can also be interpreted as a a multi marginal Optimal Mass Transportation with infinitely many marginals [69].

4.3. Congested Crowd motion

The volume preserving property appears naturally in this context where motion is constrained by the density of player.

- Optimal Mass Transportation and MFG theories can be an extremely powerful tool to attack some of these problems arising from spatial economics or to design new ones. For instance, various urban/traffic planning models have been proposed by Buttazzo, Santambrogio, Carlier,[9] [28] [20]) in recent years.
- Many models from PDEs and fluid mechanics have been used to give a description of *people or vehicles moving in a congested environment*. These models have to be classified according to the dimension (1D model are mostly used for cars on traffic networks, while 2D models are most suitable for pedestrians), to the congestion effects ("soft" congestion standing for the phenomenon where high densities slow down the movement, "hard" congestion for the sudden effects when contacts occur, or a certain threshold is attained), and to the possible rationality of the agents Maury et al [59] recently developed a theory for 2D hard congestion models without rationality, first in a discrete and then in a continuous framework. This model produces a PDE that is difficult to attack with usual PDE methods, but has been successfully studied via Optimal Mass Transportation techniques again related to the JKO gradient flow paradigm.

4.4. Astrophysics

In [44] and [25], the authors show that the deterministic past history of the Universe can be uniquely reconstructed from the knowledge of the present mass density field, the latter being inferred from the 3D distribution of luminous matter, assumed to be tracing the distribution of dark matter up to a known bias. Reconstruction ceases to be unique below those scales – a few Mpc – where multi-streaming becomes significant. Above 6 Mpc/h we propose and implement an effective Monge-Ampere-Kantorovich method of unique reconstruction. At such scales the Zel'dovich approximation is well satisfied and reconstruction becomes an instance of optimal mass transportation. After discretization into N point masses one obtains an assignment problem that can be handled by effective algorithms with not more than cubic time complexity in N and reasonable CPU time requirements. Testing against N-body cosmological simulations gives over 60% of exactly reconstructed points.

4.5. Image Processing and inverse problems

The Wasserstein distance between densities is the value function of the Optimal Mass Transportation problem. This distance may be considered to have "orthogonal" properties to the widely used least square distance. It is for instance quadratic with respect to dilations and translation. On the other hand it is not very sensitive to rigid transformations, [64] is an attempts at generalizing the CFD formulation in this context. The Wasserstein distance is an interesting tool for applications where distances between signals and in particular oscillatory signals need to to computed, this is assuming one understand how to transform the information into positive densities.

- Tannenbaum and co-authors have designed several variants of the CFD numerical method and applied it to warping, morphing and registration (using the Optimal Mass Transportation map) problems in medical imaging. [76] [17]
- Gabriel Peyre and co-authors [73] have proposed an easier to compute relaxation of the Wasserstein distance (the sliced Wasserstein distance) and applied it to two image processing problems: color transfer and texture mixing.
- Froese Engquist [40] use a Monge-Ampère Solver to compute the Wasserstein distance between synthetic 2D Seismic signals (After some transformations). Applications to waveform inversion and registration are discussed and simple numerical examples are presented.

4.6. Meteorology and Fluid models

In, [22] Brenier reviews in a unified framework the connection between optimal transport theory and classical convection theory for geophysical flows. Inspired by the numerical model proposed in [17], the starting point is a generalization of the Darcy-Boussinesq equations, which is a degenerate version of the Navier-Stokes-Boussinesq (NSB) equations. In a unified framework, he relates different variants of the NSB equations (in particular what he calls the generalized hydrostatic-Boussinesq equations) to various models involving optimal transport and the related Monge-Ampère equation. This includes the 2D semi-geostrophic equations [51] [38] [37] [4] [57] and some fully nonlinear versions of the so-called high-field limit of the Vlasov-Poisson system [65] and of the Keller-Segel system for chemotaxis [53] [33] .

4.7. Mesh motion/Lagragian methods

The necessity to preserve areas/volumes is a intrinsic feature of mesh deformations more generally Lagrangian numerical methods. Numerical method of Optimal Mass Transportation which preserve some notions of convexity and as a consequence the monotonicity of the computed transport maps can play a role in this context, see for instance [32] [35] [56].

4.8. Density Functionnal Theory (DFT)

The precise modeling of electron correlations continues to constitute the major obstacle in developing high-accuracy, low-cost methods for electronic structure computations in molecules and solids. The article [36] sheds a new light on the longstanding problem of how to accurately incorporate electron correlation into DFT, by deriving and analyzing the semiclassical limit of the exact Hohenberg-Kohn functional with the single-particle density ρ held fixed. In this limit, in the case of two electrons, the exact functional reduces to a very interesting functional that depends on an optimal transport map M associated with a given density ρ. The limit problem is known in the DFT literature with the optimal transport map being called a correlation function or a co-motion function , but it has not been rigorously derived, and it appears that it has not previously been interpreted as an optimal transport problem. The article [36] thereby links for the first time DFT, which is a large and very active research area in physics and chemistry, to optimal transportation theory, which has recently become a very active area in mathematics. Numerics are still widely open [26].

MUTANT Project-Team

4. Application Domains

4.1. Authoring and Performing Interactive Music

../../../../projets/mutant/IMG/ascograph_mainshot.jpg

Figure 3. Screenshot of Ascograph, the Antescofo graphical score editor

The combination of both realtime machine listening systems and reactive programming paradigms has enabled the *authoring* of interactive music systems as well as their realtime performance within a coherent synchronous framework called Antescofo. The module, developed since 2008 by the team members, has gained increasing attention within the user community worldwide with more than 40 prestigious public performances yearly. The outcomes of the teams's research will enhance the interactive and reactive aspects of this emerging paradigm as well as creating novel authoring tool for such purposes. The *AscoGraph* authoring environment developed in 2013 and shown in Figure 3 is the first step towards such authoring environments. The outcome of the ANR Project INEDIT (with LABRI and GRAME and coordinated by team leader), will further extend the use-cases of *Antescofo* for interactive multimedia pieces with more complex temporal structures and computational paradigms.

4.2. Music Post-Production.

Outcomes of our recognition and alignment paradigms can improve and ease existing workflows employed by audio engineers for mixing and editing using commercial Digital Audio Workstations (DAW) in post-production. We have initiated collaborations with audio engineers at Ircam and Paris Superior Music Conservatory (CNSMDP) to define the framework [9] and we will continue to develop and integrate our tools into their daily workflow.

4.3. Realtime Music Information Retrieval

We are considering to apply our information geometric approach to well-known and complex MIR problems. A glance of such problems is presented in [6]. Such applications can be used as front-end of many high-level MIR applications such as audio summarisation, audio finger printing, and automatic annotation tools. Besides such low-level enhancements, our information geometric approach can address the well-known (and still to be solved) problem of audio queries over a database.

4.4. Automatic Accompaniment/Creative Tools for Entertainment Industry

Figure 4. Automatic Accompaniment Session with Antescofo during ACM CHI 2013 Conference

Technologies developed by MuTant can find their way with general public (besides professional musicians) and within the entertainment industry. Recent trends in music industry show signs of tendencies towards more intelligent and interactive interfaces for music applications. Among them is reactive and adaptive automatic accompaniment and performance assessment as commercialized by companies such as *MakeMusic* and *Tonara*. Technologies developed around *Antescofo* can enhance interaction between user and the computer for such large public applications. We hope to pursue this by licensing our technologies to third-party companies.

PARKAS Project-Team

4. Application Domains

4.1. Domain

The project addresses the design, semantics and implementation of programming languages together with compilation techniques to develop provably safe and efficient computing systems. Traditional applications can be found in safety critical embedded systems with hard real-time constraints such as avionics (e.g., fly-by-wire command), railways (e.g., on board control, engine control), nuclear plants (e.g., emergency control of the plant). While embedded applications have been centralized, they are now massively parallel and physically distributed (e.g., sensor networks, train tracking, distributed simulation of factories) and they integrate computationally intensive algorithms (e.g., video processing) with a mix of hard and soft real-time constraints. Finally, systems are heterogeneous with discrete devices communicating with physical ones (e.g., interface between analog and digital circuits). Programming and simulating a whole system from a unique source code, with static guarantees on the reproducibility of simulations together with a compiler to generate target embedded code is a scientific and industrial challenge of great importance.

PI.R2 Project-Team (section vide)

POLSYS Project-Team

4. Application Domains

4.1. Cryptology

We propose to develop a systematic use of structured systems in Algebraic Cryptanalysis. We want to improve the efficiency and to predict the theoretical complexity of such attacks. We plan to demonstrate the power of algebraic techniques in new areas of cryptography such as Algebraic Number Theory (typically, in curve based cryptography).

4.2. Engineering sciences

Solving polynomial systems over the reals arise as a critical issue in wide range of problems coming from engineering sciences (biology, physics, control theory, etc.). We will focus on developing general enough software that may impact on these domains with a particular focus on control theory

POMDAPI Project-Team

3. Application Domains

3.1. Environmental sciences

Applications are in hydrogeology and water resources.

3.2. Energy sciences

Applications are in oil reservoir and sedimentary basin simulations, and in optimization of the power flow in an electricity transportation network.

PROSECCO Project-Team

4. Application Domains

4.1. Cryptographic protocol implementations

Cryptographic protocols such as TLS, SSH, IPSec, and Kerberos are the trusted base on which the security of modern distributed systems is built. Our work enables the analysis and verification of such protocols, both in their design and implementation. Hence, for example, we build and verify models and reference implementations for well-known protocols such as TLS, as well as analyze their popular implementations such as OpenSSL.

4.2. Hardware-based security APIs

Cryptographic devices such as Hardware Security Modules (HSMs) and smartcards are used to protect long-terms secrets in tamper-proof hardware, so that even attackers who gain physical access to the device cannot obtain its secrets. These devices are used in a variety of scenarios ranging from bank servers to transportation cards (e.g. Navigo). Our work investigates the security of commercial cryptographic hardware and evaluates the APIs they seek to implement.

4.3. Web application security

Web applications use a variety of cryptographic techniques to securely store and exchange sensitive data for their users. For example, a website may authenticate authorize users using a single sign-on protocol such as OAuth, a cloud storage service may encrypt user files on the server-side using XML encryption, and a password manager may encrypt passwords in the browser using a JavaScript cryptographic library. We build verification tools that can analyze such usages in commercial web applicaitons and evaluate their security against sophisticated web-based attacks.

RAP Project-Team (section vide)

REGAL Project-Team (section vide)

REO Project-Team

4. Application Domains

4.1. Blood flows

Cardiovascular diseases like atherosclerosis or aneurysms are a major cause of mortality. It is generally admitted that a better knowledge of local flow patterns could improve the treatment of these pathologies (although many other biophysical phenomena obviously take place in the development of such diseases). In particular, it has been known for years that the association of low wall shear stress and high oscillatory shear index give relevant indications to localize possible zones of atherosclerosis. It is also known that medical devices (graft or stent) perturb blood flows and may create local stresses favorable with atherogenesis. Numerical simulations of blood flows can give access to this local quantities and may therefore help to design new medical devices with less negative impacts. In the case of aneurysms, numerical simulations may help to predict possible zones of rupture and could therefore give a guide for treatment planning.

In clinical routine, many indices are used for diagnosis. For example, the size of a stenosis is estimated by a few measures of flow rate around the stenosis and by application of simple fluid mechanics rules. In some situations, for example in the case a sub-valvular stenosis, it is known that such indices often give false estimations. Numerical simulations may give indications to define new indices, simple enough to be used in clinical exams, but more precise than those currently used.

It is well-known that the arterial circulation and the heart (or more specifically the left ventricle) are strongly coupled. Modifications of arterial walls or blood flows may indeed affect the mechanical properties of the left ventricle. Numerical simulations of the arterial tree coupled to the heart model could shed light on this complex relationship.

One of the goals of the REO team is to provide various models and simulation tools of the cardiovascular system. The scaling of these models will be adapted to the application in mind: low resolution for modeling the global circulation, high resolution for modeling a small portion of vessel.

4.2. Respiratory tracts

Breathing, or “external” respiration (“internal” respiration corresponds to cellular respiration) involves gas transport though the respiratory tract with its visible ends, nose and mouth. Air streams then from the pharynx down to the trachea. Food and drink entry into the trachea is usually prevented by the larynx structure (epiglottis). The trachea extends from the neck into the thorax, where it divides into right and left main bronchi, which enter the corresponding lungs (the left being smaller to accommodate the heart). Inhaled air is then convected in the bronchus tree which ends in alveoli, where gaseous exchange occurs. Surfactant reduces the surface tension on the alveolus wall, allowing them to expand. Gaseous exchange relies on simple diffusion on a large surface area over a short path between the alveolus and the blood capillary under concentration gradients between alveolar air and blood. The lungs are divided into lobes (three on the right, two on the left) supplied by lobar bronchi. Each lobe of the lung is further divided into segments (ten segments of the right lung and eight of the left). Inhaled air contains dust and debris, which must be filtered, if possible, before they reach the alveoli. The tracheobronchial tree is lined by a layer of sticky mucus, secreted by the epithelium. Particles which hit the side wall of the tract are trapped in this mucus. Cilia on the epithelial cells move the mucous continually towards the nose and mouth.

Each lung is enclosed in a space bounded below by the diaphragm and laterally by the chest wall and the mediastinum. The air movement is achieved by alternately increasing and decreasing the chest pressure (and volume). When the airspace transmural pressure rises, air is sucked in. When it decreases, airspaces collapse and air is expelled. Each lung is surrounded by a pleural cavity, except at its hilum where the inner pleura give birth to the outer pleura. The pleural layers slide over each other. The tidal volume is nearly equal to 500 ml.

The lungs may fail to maintain an adequate supply of air. In premature infants surfactant is not yet active. Accidental inhalation of liquid or solid and airway infection may occur. Chronic obstructive lung diseases and lung cancers are frequent pathologies and among the three first death causes in France.

One of the goals of REO team in the ventilation field is to visualize the airways (virtual endoscopy) and simulate flow in image-based 3D models of the upper airways (nose, pharynx, larynx) and the first generations of the tracheobronchial tree (trachea is generation 0), whereas simple models of the small bronchi and alveoli are used (reduced-basis element method, fractal homogenization, multiphysics homogenization, lumped parameter models), in order to provide the flow distribution within the lung segments. This activity has been carried out in the framework of successive research programs: RNTS “R-MOD” until 2005, ACI “le-poumon-vous-dis-je” until 2007 and ANR M3RS until 2013.

4.3. Cardiac electrophysiology

The purpose is to simulate the propagation of the action potential in the heart. A lot of works has already been devoted to this topic in the literature (see *e.g.* [77], [82], [81] and the references therein), nevertheless there are only very few studies showing realistic electrocardiograms obtained from partial differential equations models. Our goal is to find a compromise between two opposite requirements: on the one hand, we want to use predictive models, and therefore models based on physiology, on the other hand, we want to use models simple enough to be parametrized (in view of patient-specific simulations). We are now working on using our ECG simulator to address the inverse problem of electrocardiology. In collaboration with the Macsproject-team, we are working on the electromechanical coupling in the myocardium. We are also interested in various clinical and industrial issues related to cardiac electrophysiology. In particular, we collaborated with ELA Medical company (pacemaker manufacturer, Sorin group).

SECRET Project-Team

4. Application Domains

4.1. Domain

Our main application domains are:

- cryptology, including classical cryptology and quantum cryptography,
- error-correcting codes, especially codes for quantum communications and fault-tolerant quantum computing,
- reverse-engineering of communication systems.

We also investigate some cross-disciplinary domains, which require a scientific competence coming from other areas, mainly social aspects of cryptology, cryptology for large databases.

SIERRA Project-Team

4. Application Domains

4.1. Application Domains

Machine learning research can be conducted from two main perspectives: the first one, which has been dominant in the last 30 years, is to design learning algorithms and theories which are as generic as possible, the goal being to make as few assumptions as possible regarding the problems to be solved and to let data speak for themselves. This has led to many interesting methodological developments and successful applications. However, we believe that this strategy has reached its limit for many application domains, such as computer vision, bioinformatics, neuro-imaging, text and audio processing, which leads to the second perspective our team is built on: Research in machine learning theory and algorithms should be driven by interdisciplinary collaborations, so that specific prior knowledge may be properly introduced into the learning process, in particular with the following fields:

- Computer vision: objet recognition, object detection, image segmentation, image/video processing, computational photography. In collaboration with the Willow project-team.
- Bioinformatics: cancer diagnosis, protein function prediction, virtual screening. In collaboration with Institut Curie.
- Text processing: document collection modeling, language models.
- Audio processing: source separation, speech/music processing. In collaboration with Telecom Paristech.
- Neuro-imaging: brain-computer interface (fMRI, EEG, MEG). In collaboration with the Parietal project-team.

4.2. Natural Language Processing

This year, our research has focused on new application domains within natural language processing (NLP), with our first two publications in leading conferences in NLP. We have worked on large-scale semantic role labelling (E. Grave, F. Bach, G. Obozinski), where we use syntactic dependency trees and learned representations from large corpora (e.g., 14.7 millions sentences, 310 millions tokens). We also extended our original work on structured sparsity to language models (F. Bach, A. Nelakanti, in collaboration with Xerox), in order to predict a word given *all* previous words, with a potentially infinite feature space organized with structured regularization.

SISYPHE Project-Team

4. Application Domains

4.1. Mathematical neuroendocrinology

Mathematical neuroendocrinology is a new field that uses mathematical modeling and analysis to help interpret neuroendocrine knowledge and design new functional assumptions or experiments. Neuroendocrinology itself is a biological scientific field at the interface between Neurosciences, Endocrinology and Physiology (and even of Developmental Biology in the case of the HPG axis) ; it studies neural networks in the brain that regulate, and that form, neuroendocrine systems.

Neuroendocrinology necessarily includes the understanding and study of peripheral physiological systems that are regulated by neuroendocrine mechanisms. Hence, in addition to our studies dedicated to the hypothalamic and pituitary levels, we do embed the target peripheral system (the gonads) in our approach of the HPG axis, with a special interest in the cell dynamics processes involved in the morphogenesis of ovarian follicles.

On the central level, we are specifically interested in the following crucial questions arising from basic and clinical neuroendocrinology: (i) How does the network-level superslow secretion rhythm of the hypothalamic hormone GnRH emerge as pulses from the fast individual dynamics of neurons? (ii) How is GnRH pulsatility switched either on or off along the different steps of the reproductive life? (iii) How is the frequency of GnRH pulses encoded and decoded by its target pituitary cells? On the peripheral level, we address the following crucial questions arising from basic and clinical reproductive and developmental biology: (i) What are the multiscale bases of the selection process operated amongst ovarian follicles that guarantees the species-specific ovulation rate in mammals ? (ii) Which configurations of the HPG axis allow for selection escape and poly-ovulating strategies, as observed naturally in prolific species or in strain-specific genetic mutations? (iii) How does the interaction between the oocyte and its surrounding follicular cells shape the morphology of the follicle in the early stages?

4.2. Quantum engineering

A new field of quantum systems engineering has emerged during the last few decades. This field englobes a wide range of applications including nano-electro-mechanical devices, nuclear magnetic resonance applications, quantum chemical synthesis, high resolution measurement devices and finally quantum information processing devices for implementing quantum computation and quantum communication. Recent theoretical and experimental achievements have shown that the quantum dynamics can be studied within the framework of estimation and control theory, but give rise to new models that have not been fully explored yet.

The QUANTIC team's activities are defined at the theoretical and experimental border of this emerging field with an emphasis on the applications in quantum information, computation and communication. The main objective of this interdisciplinary team formed by applied mathematicians (Mazyar Mirrahimi and Pierre Rouchon) and experimental physicists (Benjamin Huard and François Mallet) is to develop quantum devices ensuring a robust processing of quantum information.

On the theory side, this is done by following a system theory approach: we develop estimation and control tools adapted to particular features of quantum systems. The most important features, requiring the development of new engineering methods, are related to the concept of measurement and feedback for composite quantum systems. The destructive and partial nature of measurements for quantum systems lead to major difficulties in extending classical control theory tools. Indeed, design of appropriate measurement protocols and, in the sequel, the corresponding quantum filters estimating the state of the system from the partial measurement record, are themselves bricks of the quantum system theory to be developed.

On the experimental side, we develop new quantum information processing devices based on quantum superconducting circuits. Indeed, by combining superconducting circuits in low temperatures and using techniques from micro-wave measurements, the macroscopic and collective degrees of freedom such as the voltage and the current are forced to behave according to the laws of quantum mechanics. Our quantum devices are aimed to protect and process the quantum information through these integrated circuits.

4.3. Monitoring and control of complex systems

Questions of modeling, identification, signal analysis and control are important in many medical or general engineering applications. We consider some very prospective questions as well as engineering questions raised by challenging industrial projects. The topics considered are the following:

Modeling, signal analysis and control with medical applications:

- *3D cardiac modeling for personalized medicine.* Our main contribution to Inria collective effort in this field (project-teams Asclepios, MACS, REO, Sisyphe) is the so-called “Bestel-Clément-Sorine” model of contraction of cardiac muscle [86], at the origin of the 3D electromechanical direct and inverse modeling of the heart at Inria. This model is based on ideas originating from the kinetic equation theory, used to model, on the molecular scale, the controlled collective behavior of actin-myosin nanomotors at the root of muscle contraction. The classical Huxley’s model was recovered on the sarcomere scale by using moment equations and a controlled constitutive law on the tissue scale was obtained using the same type of scaling techniques. The model, now embedded in heart simulators is used in various studies [55], [3], [112], [110].

- *Semiclassical analysis of cardiovascular signals.* This work began with the article [91] and the PhD of M. Laleg-Kirati [100], [99], [102]. The theory and a validation of a new method of blood pressure analysis are now published [51], [101].
The main idea is to consider a signal $x \to y(x)$ to analyze as the multiplication operator $\phi \to y\phi$ on some function space, and to analyze it as a potential. The signal is represented by the spectrum of an associated Schrödinger operator, combined with a semi-classical quantification: $-h^2\dfrac{d^2}{dx^2} - y(x)$ with $h > 0$ small. For signals looking as “superpositions of bumps” (e.g. the systolic pulse, the dichrotic notch for the arterial pulse pressure), this leads to some kind of nonlinear Fourier analysis [51]. The spectral parameters associated with the arterial pressure can be useful cardiovascular indices, e.g. for noninvasive blood flow estimation [101]. In the arterial pressure case, this is equivalent to approximate the traveling pressure pulse by a N-soliton solution of a Korteweg-de Vries (KdV) equation [91] and using ideas similar to the Lax pair representation of N-solitons and proof technique for the weak dispersion limit of KdV. A striking result is that an N-soliton is a very good representation of the arterial pressure waveform for values of N as small as $N = 3$. The representation of pulse-shaped signals is parcimonious, having only $2N$ parameters [113].

- *Multiscale signal analysis of cardiovascular signals:* collaboration with Julien Barral (former member of Sisyphe) and partners of the ANR project DMASC. The starting point was the common idea that "A Healthy Heart Is a Fractal Heart". We have developed a method to test the existence of scale laws in signals and applied it to RR signals: the heart rate is not always fractal or even multifractal in an Healthy Heart [19].

- *Modeling and control of CARMAT Total Artificial Heart.* This TAH has been implanted for the first time in a patient in Dec 2013. We have contributed to this industrial project since 2008 on modeling and control questions during the post-doc of Karima Djabella (now at CARMAT), Frédéric Vallais and the two-year contract for supervising Julien Bernard (CARMAT control engineer). It was an opportunity for valorizing some results on the baroreflex control [94] or heart rate variability during exercise [90].

- *Glycemic control in Intensive Care Units (ICUs):* Blood glucose is a key biological parameter in ICU since the study of van den Berghe et al [123] who demonstrated decreased mortality in surgical intensive care patients in association with tIght glycemic control (TGC), based on intensive insulin therapy. But there was only one ICU and the protocol was not formalized. Trying to decrease mortality in standard ICUs by using computer aided glycemic control is still a challenge. Previous studies have failed because of high rates of severe hypoglycaemia. The last one was NICE-SUGAR [117] with a 2% increase in mortality (death ratio from any cause within 90 days after randomization compared between control and TGC patients). In cooperation with Pierre Kalfon (Intensive Care, Hospital of Chartres) and in the framework of a CIFRE contract with a small medtech company LK2 (Tours, France), we have studied the origins of these failures and proposed more robust control algorithms tuned using a database of representative "virtual patients" [95], [96] and the PhD of A. Guerrini, [31]. A first version of the controller has been tested in a large clinical study CGAO-REA [70], [48].

- *Cardiorespiratory signal processing in ICUs:* cooperation with François Cottin (INSERM 902, Génopôle, Evry), Andry Van de Louw (Service de Réanimation Polyvalente, Centre Hospitalier Sud-Francilien, Evry) on the analysis of the effect of mechanical ventilation [118], [120], [119].

Modeling, signal analysis and control for general engineering:

Identification of nonlinear systems: from algorithms to a popular matlab toolbox:
- Identification of nonlinear systems: with Jiandong Wang (Associate Professor, Beijing University, China) [122], [121]: Block-oriented nonlinear system identification.
- Development of the Matlab System Identification ToolBox (SITB). See Section 5.1 .

Identification of transmission line characteristics: from algorithms to electronic experiments. Collaboration with CEA LIST (Lab of applied research on software-intensive technologies) and LGEP (Laboratoire de génie électrique de Paris) with Florent Loete [106] (ANR projects SEEDS, 0-DEFECT, INSCAN, SODDA).
We have extended to some networks the seminal work of Jaulent [97] for the real line: all the information contained in a measured reflection coefficient can be obtained by solving an inverse scattering problem for a system of Schrödinger or Zakharov-Shabat equations on the graph of the network, which allows one to recover the geometry of the network and some electrical characteristics for nonuniform lossless electrical star-shaped networks [26]. An efficient method to solve the associated Guelfand-Levitan-Marchenko equations has been studied and is used in the software ISTL (see Section 5.2) [61], [114], [115]. An engineering methodology based on this approach has been described [29] and some first experimental results obtained [106].

Monitoring and control of automotive depollution systems: with RENAULT (Karim Bencherif, Damiano Di Penta and PhD students): [75], [20], [85].

Oscillatory systems in Control: reduced modeling, analysis, identification and synthesis: this is the topic of a cooperation with ITA (São José dos Campos, Brazil) [33].

SMIS Project-Team

4. Application Domains

4.1. Application Domains

Our work addresses varied application domains. Typically, data management techniques on chip are required each time data-driven applications have to be embedded in ultra-light computing devices. This situation occurs for example in healthcare applications where medical folders are embedded into smart tokens (e.g., smart cards, secured USB keys), in telephony applications where personal data (address book, agenda, etc.) is embedded into cellular phones, in sensor networks where sensors log row measurements and perform local computation on them, in smart-home applications where a collection of smart appliances gather information about the occupants to provide them a personalized service, and more generally in most applications related to ambient intelligence.

Safeguarding data confidentiality has become a primary concern for citizens, administrations and companies, broadening the application domains of our work on access control policies definition and enforcement. The threat on data confidentiality is manifold: external and internal attacks on the data at rest, on the data on transit, on the data hosted in untrusted environments (e.g., Database Service Providers, Web-hosting companies) and subject to illegal usage, insidious gathering of personal data in an ambient intelligence surrounding. Hence, new access control models and security mechanisms are required to accurately declare and safely control who is granted access to which data and for which purpose.

While the application domain mentioned above is rather large, two applications are today more specifically targeted by the SMIS team. The first one deals with privacy preservation in EHR (Electronic Health Record) systems and PCEHR (Personnally Controlled EHR). We are developing technologies tackling this issue and experiment them in the field. The second application area deals with privacy preservation in the context of personal Cloud, that is personal data hosted in dedicated servers staying under the holder's control (e.g., in a personal internet box or in a home automation box).

WILLOW Project-Team

4. Application Domains

4.1. Introduction

We believe that foundational modeling work should be grounded in applications. This includes (but is not restricted to) the following high-impact domains.

4.2. Quantitative image analysis in science and humanities

We plan to apply our 3D object and scene modeling and analysis technology to image-based modeling of human skeletons and artifacts in anthropology, and large-scale site indexing, modeling, and retrieval in archaeology and cultural heritage preservation. Most existing work in this domain concentrates on image-based rendering—that is, the synthesis of good-looking pictures of artifacts and digs. We plan to focus instead on quantitative applications. We are engaged in a project involving the archaeology laboratory at ENS and focusing on image-based artifact modeling and decorative pattern retrieval in Pompeii. This effort is part of the MSR-Inria project mentioned earlier and that will be discussed further later in this report. Application of our 3D reconstruction technology is now being explored in the field of cultural heritage and archeology by the start-up Iconem, founded by Y. Ubelmann, a Willow collaborator.

4.3. Video Annotation, Interpretation, and Retrieval

Both specific and category-level object and scene recognition can be used to annotate, augment, index, and retrieve video segments in the audiovisual domain. The Video Google system developed by Sivic and Zisserman (2005) for retrieving shots containing specific objects is an early success in that area. A sample application, suggested by discussions with Institut National de l'Audiovisuel (INA) staff, is to match set photographs with actual shots in film and video archives, despite the fact that detailed timetables and/or annotations are typically not available for either medium. Automatically annotating the shots is of course also relevant for archives that may record hundreds of thousands of hours of video. Some of these applications will be pursued in our MSR-Inria project, in which INA is one of our partners.