Activity Report 2017

Section Application Domains

ALGORITHMICS, PROGRAMMING, SOFTWARE AND ARCHITECTURE

DIGITAL HEALTH, BIOLOGY AND EARTH

ANTIQUE Project-Team

4. Application Domains

4.1. Verification of safety critical embedded software

The verification of safety critical embedded software is a very important application domain for our group. First, this field requires a high confidence in software, as a bug may cause disastrous events. Thus, it offers an obvious opportunity for a strong impact. Second, such software usually have better specifications and a better design than many other families of software, hence are an easier target for developing new static analysis techniques (which can later be extended for more general, harder to cope with families of programs). This includes avionics, automotive and other transportation systems, medical systems...

For instance, the verification of avionics systems represent a very high percentage of the cost of an airplane (about 30 % of the overall airplane design cost). The state of the art development processes mainly resort to testing in order to improve the quality of software. Depending on the level of criticality of a software (at highest levels, any software failure would endanger the flight) a set of software requirements are checked with test suites. This approach is both costly (due to the sheer amount of testing that needs to be performed) and unsound (as errors may go unnoticed, if they do not arise on the test suite).

By contrast, static analysis can ensure higher software quality at a lower cost. Indeed, a static analyzer will catch all bugs of a certain kind. Moreover, a static analysis run typically lasts a few hours, and can be integrated in the development cycle in a seamless manner. For instance, ASTRÉE successfully verified the absence of runtime error in several families of safety critical fly-by-wire avionic software, in at most a day of computation, on standard hardware. Other kinds of synchronous embedded software have also been analyzed with good results.

In the future, we plan to greatly extend this work so as to verify *other families of embedded software* (such as communication, navigation and monitoring software) and *other families of properties* (such as security and liveness properties).

Embedded software in charge of communication, navigation, monitoring typically rely on a *parallel* structure, where several threads are executed in parallel, and manage different features (input, output, user interface, internal computation, logging...). This structure is also often found in automotive software. An even more complex case is that of *distributed* systems, where several separate computers are run in parallel and take care of several sub-tasks of a same feature, such as braking. Such a logical structure is not only more complex than the synchronous one, but it also introduces new risks and new families of errors (deadlocks, data-races...). Moreover, such less well designed, and more complex embedded software often utilizes more complex data-structures than synchronous programs (which typically only use arrays to store previous states) and may use dynamic memory allocation, or build dynamic structures inside static memory regions, which are actually even harder to verify than conventional dynamically allocated data structures. Complex data-structures also introduce new kinds of risks (the failure to maintain structural invariants may lead to runtime errors, non termination, or other software failures). To verify such programs, we will design additional abstract domains, and develop new static analysis techniques, in order to support the analysis of more complex programming language features such as parallel and concurrent programming with threads and manipulations of complex data structures. Due to their size and complexity, the verification of such families of embedded software is a major challenge for the research community.

Furthermore, embedded systems also give rise to novel security concerns. It is in particular the case for some aircraft-embedded computer systems, which communicate with the ground through untrusted communication media. Besides, the increasing demand for new capabilities, such as enhanced on-board connectivity, e.g. using mobile devices, together with the need for cost reduction, leads to more integrated and interconnected systems. For instance, modern aircrafts embed a large number of computer systems, from safety-critical cockpit avionics to passenger entertainment. Some systems meet both safety and security requirements.

Despite thorough segregation of subsystems and networks, some shared communication resources raise the concern of possible intrusions. Because of the size of such systems, and considering that they are evolving entities, the only economically viable alternative is to perform automatic analyses. Such analyses of security and confidentiality properties have never been achieved on large-scale systems where security properties interact with other software properties, and even the mapping between high-level models of the systems and the large software base implementing them has never been done and represents a great challenge. Our goal is to prove empirically that the security of such large scale systems can be proved formally, thanks to the design of dedicated abstract interpreters.

The long term goal is to make static analysis more widely applicable to the verification of industrial software.

4.2. Static analysis of software components and libraries

An important goal of our work is to make static analysis techniques easier to apply to wider families of software. Then, in the longer term, we hope to be able to verify less critical, yet very commonly used pieces of software. Those are typically harder to analyze than critical software, as their development process tends to be less rigorous. In particular, we will target operating systems components and libraries. As of today, the verification of such programs is considered a major challenge to the static analysis community.

As an example, most programming languages offer Application Programming Interfaces (API) providing ready-to-use abstract data structures (e.g., sets, maps, stacks, queues, etc.). These APIs, are known under the name of containers or collections, and provide off-the-shelf libraries of high level operations, such as insertion, deletion and membership checks. These container libraries give software developers a way of abstracting from low-level implementation details related to memory management, such as dynamic allocation, deletion and pointer handling or concurrency aspects, such as thread synchronization. Libraries implementing data structures are important building bricks of a huge number of applications, therefore their verification is paramount. We are interested in developing static analysis techniques that will prove automatically the correctness of large audience libraries such as Glib and Threading Building Blocks.

4.3. Biological systems

Computer Science takes a more and more important role in the design and the understanding of biological systems such as signaling pathways, self assembly systems, DNA repair mechanisms. Biology has gathered large data-bases of facts about mechanistic interactions between proteins, but struggles to draw an overall picture of how these systems work as a whole. High level languages designed in Computer Science allow to collect these interactions in integrative models, and provide formal definitions (i.e., semantics) for the behavior of these models. This way, modelers can encode their knowledge, following a bottom-up discipline, without simplifying *a priori* the models at the risk of damaging the key properties of the system. Yet, the systems that are obtained this way suffer from combinatorial explosion (in particular, in the number of different kinds of molecular components, which can arise at run-time), which prevents from a naive computation of their behavior.

We develop various abstract interpretation-based analyses, tailored to different phases of the modeling process. We propose automatic static analyses in order to detect inconsistencies in the early phases of the modeling process. These analyses are similar to the analysis of classical safety properties of programs. They involve both forward and backward reachability analyses as well as causality analyses, and can be tuned at different levels of abstraction. We also develop automatic static analyses so as to identify the key elements in the dynamics of these models. The results of these analyses are sent to another tool, which is used to automatically simplify the models. The correctness of this simplification process is proved by the means of abstract interpretation: this ensures formally that the simplification preserves the quantitative properties that have been specified beforehand by the modeler. The whole pipeline is parameterized by a large choice of abstract domains which exploits different features of the high level description of models.

<p style="text-align:center">**AOSTE2 Team**</p>

4. Application Domains

4.1. Avionics

Participants: Liliana Cucu, Keryan Didier, Adriana Gogonel, Cristian Maxim, Dumitru Potop-Butucaru, Yves Sorel.

A large number of our activities, in analysis, modelling, design and implementation of real-time embedded systems addresses specific applications mainly in the avionics field (with partners such as Airbus, Thales, Safran, etc.) (in the CAPACITES and ASSUME projects 9.1.2.1 , 9.2.1.1).

4.2. Many-Core Embedded Architectures

Participants: Liliana Cucu, Keryan Didier, Dumitru Potop-Butucaru, Anselme Revuz, Yves Sorel.

The AAA approach (fitting embedded applications onto embedded architectures) requires a sufficiently precise description of (a model of) the architecture (description platform). Such platforms become increasingly heterogeneous, and we had to consider a number of emerging ones with that goal in mind, such as Kalray MPPA (in the CAPACITES and ASSUME projects 9.1.2.1 , 9.2.1.1).

4.3. Railways

Participants: Liliana Cucu, Adriana Gogonel, Walid Talaboulma.

The statistical estimation of bounds on the execution time of a program on a processor is applied in the context of railroad crossing in the context of the collaborative project DEPARTS 9.1.2.2 .

ARIC Project-Team

4. Application Domains

4.1. Floating-point and Validated Numerics

Our expertise on validated numerics is useful to analyze and improve, and guarantee the quality of numerical results in a wide range of applications including:

- scientific simulation;
- global optimization;
- control theory.

Much of our work, in particular the development of correctly rounded elementary functions, is critical to the

- reproducibility of floating-point computations.

4.2. Cryptography, Cryptology, Communication Theory

Lattice reduction algorithms have direct applications in

- public-key cryptography;
- diophantine equations;
- communications theory.

AROMATH Project-Team

4. Application Domains

4.1. Geometric modeling for Design and Manufacturing.

The main domain of applications that we consider for the methods we develop is Computer Aided Design and Manufacturing.

Computer-Aided Design (CAD) involves creating digital models defined by mathematical constructions, from geometric, functional or aesthetic considerations. Computer-aided manufacturing (CAM) uses the geometrical design data to control the tools and processes, which lead to the production of real objects from their numerical descriptions.

CAD-CAM systems provide tools for visualizing, understanding, manipulating, and editing virtual shapes. They are extensively used in many applications, including automotive, shipbuilding, aerospace industries, industrial and architectural design, prosthetics, and many more. They are also widely used to produce computer animation for special effects in movies, advertising and technical manuals, or for digital content creation. Their economic importance is enormous. Their importance in education is also growing, as they are more and more used in schools and educational purposes.

CAD-CAM has been a major driving force for research developments in geometric modeling, which leads to very large software, produced and sold by big companies, capable of assisting engineers in all the steps from design to manufacturing.

Nevertheless, many challenges still need to be addressed. Many problems remain open, related to the use of efficient shape representations, of geometric models specific to some application domains, such as in architecture, naval engineering, mechanical constructions, manufacturing .... Important questions on the robustness and the certification of geometric computation are not yet answered. The complexity of the models which are used nowadays also appeals for the development of new approaches. The manufacturing environment is also increasingly complex, with new type of machine tools including: turning, 5-axis machining and wire EDM (Electrical Discharge Machining), 3D printer. It cannot be properly used without computer assistance, which raises methodological and algorithmic questions. There is an increasing need to combine design and simulation, for analyzing the physical behavior of a model and for optimal design.

The field has deeply changed over the last decades, with the emergence of new geometric modeling tools built on dedicated packages, which are mixing different scientific areas to address specific applications. It is providing new opportunities to apply new geometric modeling methods, output from research activities.

4.2. Geometric modeling for Numerical Simulation and Optimization

A major bottleneck in the CAD-CAM developments is the lack of interoperability of modeling systems and simulation systems. This is strongly influenced by their development history, as they have been following different paths.

The geometric tools have evolved from supporting a limited number of tasks at separate stages in product development and manufacturing, to being essential in all phases from initial design through manufacturing.

Current Finite Element Analysis (FEA) technology was already well established 40 years ago, when CAD-systems just started to appear, and its success stems from using approximations of both the geometry and the analysis model with low order finite elements (most often of degree ≤ 2).

There has been no requirement between CAD and numerical simulation, based on Finite Element Analysis, leading to incompatible mathematical representations in CAD and FEA. This incompatibility makes interoperability of CAD/CAM and FEA very challenging. In the general case today this challenge is addressed by expensive and time-consuming human intervention and software developments.

Improving this interaction by using adequate geometric and functional descriptions should boost the interaction between numerical analysis and geometric modeling, with important implications in shape optimization. In particular, it could provide a better feedback of numerical simulations on the geometric model in a design optimization loop, which incorporates iterative analysis steps.

The situation is evolving. In the past decade, a new paradigm has emerged to replace the traditional Finite Elements by B-Spline basis element of any polynomial degree, thus in principle enabling exact representation of all shapes that can be modeled in CAD. It has been demonstrated that the so-called isogeometric analysis approach can be far more accurate than traditional FEA.

It opens new perspectives for the interoperability between geometric modeling and numerical simulation. The development of numerical methods of high order using a precise description of the shapes raises questions on piecewise polynomial elements, on the description of computational domains and of their interfaces, on the construction of good function spaces to approximate physical solutions. All these problems involve geometric considerations and are closely related to the theory of splines and to the geometric methods we are investigating. We plan to apply our work to the development of new interactions between geometric modeling and numerical solvers.

<div style="text-align:center; color:red; font-weight:bold;">CAIRN Project-Team</div>

4. Application Domains

4.1. Panorama

keywords: Wireless (Body) Sensor Networks, High-Rate Optical Communications, Wireless Communications, Applied Cryptography.

Our research is based on realistic applications, in order to both discover the main needs created by these applications and to invent realistic and interesting solutions.

Wireless Communication is our privileged application domain. Our research includes the prototyping of (subsets of) such applications on reconfigurable and programmable platforms. For this application domain, the high computational complexity of the 5G Wireless Communication Systems calls for the design of high-performance and energy-efficient architectures. In **Wireless Sensor Networks** (WSN), where each wireless node is expected to operate without battery replacement for significant periods of time, energy consumption is the most important constraint. Sensor networks are a very dynamic domain of research due, on the one hand, to the opportunity to develop innovative applications that are linked to a specific environment, and on the other hand to the challenge of designing totally autonomous communicating objects.

Other important fields are also considered: hardware cryptographic and security modules, high-rate optical communications, machine learning, and multimedia processing.

<div align="center">**CAMUS Team**</div>

4. Application Domains

4.1. Application Domains

Performance being our main objective, our developments' target applications are characterized by intensive computation phases. Such applications are numerous in the domains of scientific computations, optimization, data mining and multimedia.

Applications involving intensive computations are necessarily high energy consumers. However this consumption can be significantly reduced thanks to optimization and parallelization. Although this issue is not our primary objective, we can expect some positive effects for the following reasons:

- Program parallelization tries to distribute the workload equally among the cores. Thus an equivalent performance, or even a better performance, to a sequential higher frequency execution on one single core, can be obtained.

- Memory and memory accesses are high energy consumers. Lowering the memory consumption, lowering the number of memory accesses and maximizing the number of accesses in the low levels of the memory hierarchy (registers, cache memories) have a positive consequence on execution speed, but also on energy consumption.

CARAMBA Project-Team

4. Application Domains

4.1. Better Awareness and Avoidance of Cryptanalytic Threats

Our study of the Number Field Sieve family of algorithms aims at showing how the threats underlying various supposedly hard problems are real. Our record computations, as well as new algorithms, contribute to having a scientifically accurate assessment of the feasibility limit for these problems, given academic computing resources. The data we provide in this way is a primary ingredient for government agencies whose purpose includes guidance for the choice of appropriate cryptographic primitives. For example the French ANSSI [0], German BSI, or the NIST [0] in the United States base their recommendations on such computational achievements.

The software we make available to achieve these cryptanalytic computations also allows us to give cost estimates for potential attacks to cryptographic systems that are taking the security/efficiency/legacy compatibility trade-offs too lightly. Attacks such as LogJam [24] are understood as being serious concerns thanks to our convincing proof-of-concepts. In the LogJam context, this impact has led to rapid worldwide security advisories and software updates that eventually defeat some potential intelligence threats and improve confidentiality of communications.

4.2. Promotion of Better Cryptography

We also promote the switch to algebraic curves as cryptographic primitives. Those offer nice speed and excellent security, while primitives based on elementary number theory (integer factorization, discrete logarithm in finite fields), which underpin e.g., RSA, are gradually forced to adopt unwieldy key sizes so as to comply with the desired security guarantees of modern cryptography. Our contributions to the ultimate goal of having algebraic curves eventually take over the cryptographic landscape lie in our fast arithmetic contributions, our contributions to the point counting problem, and more generally our expertise on the diverse surrounding mathematical objects, or on the special cases where the discrete logarithm problem is not hard enough and should be avoided.

We also promote cryptographically sound electronic voting, for which we develop the Belenios prototype software, (licensed under the AGPL). It depends on research made in collaboration with the PESTO team, and provides stronger guarantees than current state of the art.

4.3. Key Software Tools

The vast majority of our work is eventually realized as software. We can roughly categorize it in two groups. Some of our software covers truly fundamental objects, such as the GNU MPFR, GNU MPC, GF2X, or MPFQ packages. To their respective extent, these software packages are meant to be included or used in broader projects. For this reason, it is important that the license chosen for this software allows proper reuse, and we favor licenses such as the LGPL, which is not restrictive. We can measure the impact of this software by the way it is used in e.g., the GNU Compiler Collection (GCC), in Victor Shoup's Number Theory Library (NTL), or in the Sage computer algebra system. The availability of these software packages in most Linux distributions is also a good measure for the impact of our work.

---

[0]In [25], the minimal recommended RSA key size is 2048 bits for usage up to 2030. See also Annex B, in particular Section B.1 "Records de calculs cryptographiques".

[0]The work [36] is one of only two academic works cited by NIST in the initial version (2011) of the report [40].

We also develop more specialized software. Our flagship software package is Cado-NFS, and we also develop some others with various levels of maturity, such as GMP-ECM, CMH, or Belenios, aiming at quite diverse targets. Within the lifespan of the CARAMBA project, we expect more software packages of this kind to be developed, specialized towards tasks relevant to our research targets: important mathematical structures attached to genus 2 curves, generation of cryptographically secure curves, or tools for attacking cryptographically hard problems. Such software both illustrates our algorithms, and provides a base on which further research work can be established. Because of the very nature of these specialized software packages as research topics in their own right, needing both to borrow material from other projects, and being possible source of inspiring material for others, it is again important that these be developed in a free and open-source development model.

<p style="text-align:center"><strong style="color:red">CARTE Team</p>

4. Application Domains

4.1. Continuous computation theories

Understanding computation theories for continuous systems leads to studying hardness of verification and control of these systems. This has been used to discuss problems in fields as diverse as verification (see e.g., [39]), control theory (see e.g., [47]), neural networks (see e.g., [72]), and so on. We are interested in the formal decidability of properties of dynamical systems, such as reachability [63], the Skolem-Pisot problem [44], the computability of the ω-limit set [62]. Those problems are analogous to verification of safety properties.

Contrary to computability theory, complexity theory over continuous spaces is underdeveloped and not well understood. A central issue is the choice of the representation of objects by discrete data and its effects on the induced complexity notions. As for computability, it is well known that a representation is gauged by the topology it induces. However more structure is needed to capture the complexity notions: topologically equivalent representations may induce different classes of polynomial-time computable objects, e.g., developing a sound complexity theory over continuous structures would enable us to make abstract computability results more applicable by analyzing the corresponding complexity issues. We think that the preliminary step towards such a theory is the development of higher-order complexity, which we are currently carrying out.

In contrast with the discrete setting, it is of utmost importance to compare the various models of computation over the reals, as well as their associated complexity theories. In particular, we focus on the General Purpose Analog Computer of Claude Shannon [75], on recursive analysis [79], on the algebraic approach [70] and on Markov computability [64]. A crucial point for future investigations is to fill the gap between continuous and discrete computational models. This is one deep motivation of our work on computation theories for continuous systems.

4.2. Analysis and verification of adversary systems

The other research direction on dynamical systems we are interested in is the study of properties of adversary systems or programs, i.e., of systems whose behavior is unknown or indistinct, or which do not have classical expected properties. We would like to offer proof and verification tools, to guarantee the correctness of such systems. On one hand, we are interested in continuous and hybrid systems. In a mathematical sense, a hybrid system can be seen as a dynamical system, whose transition function does not satisfy the classical regularity hypotheses, like continuity, or continuity of its derivative. The properties to be verified are often expressed as reachability properties. For example, a safety property is often equivalent to (non-)reachability of a subset of unsure states from an initial configuration, or to stability (with its numerous variants like asymptotic stability, local stability, mortality, etc.). Thus we will essentially focus on verification of these properties in various classes of dynamical systems.

We are also interested in rewriting techniques, used to describe dynamic systems, in particular in the adversary context. As they were initially developed in the context of automated deduction, the rewriting proof techniques, although now numerous, are not yet adapted to the complex framework of modelling and programming. An important stake in the domain is then to enrich them to provide realistic validation tools, both in providing finer rewriting formalisms and their associated proof techniques, and in developing new validation concepts in the adversary case, i.e., when usual properties of the systems like, for example, termination are not verified. For several years, we have been developing specific procedures for property proofs of rewriting, for the sake of programming, in particular with an inductive technique, already applied with success to termination under strategies [54], [55], [56], to weak termination [57], sufficient completeness [58] and probabilistic termination [60]. The last three results take place in the context of adversary computations, since they allow for proving that

even a divergent program, in the sense where it does not terminate, can give the expected results. A common mechanism has been extracted from the above works, providing a generic inductive proof framework for properties of reduction relations, which can be parametrized by the property to be proved [59], [61]. Provided program code can be translated into rule-based specifications, this approach can be applied to correctness proofs of software in a larger context. A crucial element of safety and security of software systems is the problem of resources. We are working in the field of Implicit Computational Complexity. Interpretation based methods like Quasi-interpretations (QI) or sup-interpretations, are the approach we have been developing these last years [66], [67], [68]. Implicit complexity is an approach to the analysis of the resources that are used by a program. Its tools come essentially from proof theory. The aim is to compile a program while certifying its complexity.

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.

CELTIQUE Project-Team (section vide)

<p style="text-align:center;color:red;font-weight:bold">CIDRE Project-Team</p>

4. Application Domains

4.1. Security is Required Everywhere

With the infiltration of computers and software in almost all aspects of our modern life, security can nowadays be seen as an absolutely general concern. As such, the results of the research targeted by CIDRE apply to a wide range of domains. It is clear that critical systems, in which security (and safety) is a major concern, can benefit from ideas such as dynamic security policy monitoring. On the other hand, systems used by the general public (basically, the internet and services such as web or cloud services, social networks, location-based services, etc.) can also benefit from results obtained by CIDRE, in particular to solve some of the privacy issues raised by these systems that manipulate huge amount of personal data. In addition, systems are getting more and more complex, decentralized, distributed, or spontaneous. Cloud computing, in particular, brings many challenges that could benefit from ideas, approaches and solutions studied by CIDRE in the context of distributed systems.

Industrial Control Systems (ICS) and in particular Supervisory Control and Data Acquisition are also new application domains for intrusion detection. The Stuxnet attack has emphasized the vulnerability of such critical systems which are not totally isolated anymore. Securing ICS is challenging since modifications of the systems, for example to patch them, are often not possible. High availability requirements also often conflict with preventive approaches. In this case, security monitoring is appealing to protect such systems against malicious activities. Intrusion detection in ICS is not fundamentally different from traditional approaches. However, new hypotheses and constraints need to be taken into account, which also bring interesting new research challenges.

COMETE Project-Team

4. Application Domains

4.1. Security and privacy

Participants: Konstantinos Chatzikokolakis, Catuscia Palamidessi, Ali Kassem, Anna Pazii, Tymofii Prokopenko.

The aim of our research is the specification and verification of protocols used in mobile distributed systems, in particular security protocols. We are especially interested in protocols for *information hiding*.

Information hiding is a generic term which we use here to refer to the problem of preventing the disclosure of information which is supposed to be secret or confidential. The most prominent research areas which are concerned with this problem are those of *secure information flow* and of *privacy*.

Secure information flow refers to the problem of avoiding the so-called *propagation* of secret data due to their processing. It was initially considered as related to software, and the research focussed on type systems and other kind of static analysis to prevent dangerous operations, Nowadays the setting is more general, and a large part of the research effort is directed towards the investigation of probabilistic scenarios and treaths.

Privacy denotes the issue of preventing certain information to become publicly known. It may refer to the protection of *private data* (credit card number, personal info etc.), of the agent's identity (*anonymity*), of the link between information and user (*unlinkability*), of its activities (*unobservability*), and of its *mobility* (*untraceability*).

The common denominator of this class of problems is that an adversary can try to infer the private information (*secrets*) from the information that he can access (*observables*). The solution is then to obfuscate the link between secrets and observables as much as possible, and often the use randomization, i.e. the introduction of *noise*, can help to achieve this purpose. The system can then be seen as a *noisy channel*, in the information-theoretic sense, between the secrets and the observables.

We intend to explore the rich set of concepts and techniques in the fields of information theory and hypothesis testing to establish the foundations of quantitive information flow and of privacy, and to develop heuristics and methods to improve mechanisms for the protection of secret information. Our approach will be based on the specification of protocols in the probabilistic asynchronous π-calculus, and the application of model-checking to compute the matrices associated to the corresponding channels.

CONVECS Project-Team

4. Application Domains

4.1. Application Domains

The theoretical framework we use (automata, process algebras, bisimulations, temporal logics, etc.) and the software tools we develop are general enough to fit the needs of many application domains. They are applicable to virtually any system or protocol that consists of distributed agents communicating by asynchronous messages. The list of recent case studies performed with the CADP toolbox (see in particular § 6.5) illustrates the diversity of applications:

- *Bioinformatics:* genetic regulatory networks, nutritional stress response, metabolic pathways,
- *Component-based systems:* Web services, peer-to-peer networks,
- *Databases:* transaction protocols, distributed knowledge bases, stock management,
- *Distributed systems:* virtual shared memory, dynamic reconfiguration algorithms, fault tolerance algorithms, cloud computing,
- *Embedded systems:* air traffic control, avionic systems, medical devices,
- *Hardware architectures:* multiprocessor architectures, systems on chip, cache coherency protocols, hardware/software codesign,
- *Human-machine interaction:* graphical interfaces, biomedical data visualization, plasticity,
- *Security protocols:* authentication, electronic transactions, cryptographic key distribution,
- *Telecommunications:* high-speed networks, network management, mobile telephony, feature interaction detection.

<p align="center" style="color:red">CORSE Project-Team</p>

4. Application Domains

4.1. Transfer

The main industrial sector related to the research activities of CORSE is the one of semi-conductor (programmable architectures spanning from embedded systems to servers). Obviously any computing application which has the objective of exploiting as much as possible the resources (in terms of high-performance but also low energy consumption) of the host architecture is intended to take advantage of advances in compiler and run-time technology. These applications are based over numerical kernels (linear algebra, FFT, convolution...) that can be adapted on a large spectrum of architectures. Members of CORSE already maintain fruitful and strong collaborations with several companies such as STMicroelectronics, Bull, UpMem, Kalray, or Aselta.

DATASHAPE Project-Team

4. Application Domains

4.1. Main application domains

Our work is mostly of a fundamental mathematical and algorithmic nature but fin ds applications in a variety of application in data analysis, more precisely in Topological Data Analysis (TDA). Although TDA is a quite recent field, it al ready founds applications in material science, biology, sensor networks, 3D shap es analysis and processing, to name a few.

More specifically, DATASHAPEhas recently started to work on the analysis of t rajectories obtained from inertial sensors (starting PhD thesis of Bertrand Bea ufils) and is exploring some possible new applications in material science.

DATASPHERE Team

4. Application Domains

4.1. Governance

Application domains include:

- City governance, data and interaction with citizens.
- New voting mechanisms.
- Local governance vs global norms and control.
- Security in an unstable world.
- Adaptation to the conditions of the anthropocene, resources and homeostasis.
- Data strategy for digital economy

<p style="text-align:center">**DEDUCTEAM Project-Team**</p>

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. Termination certificate verification

Termination is an important property to verify, especially in critical applications. Automated termination provers use more and more complex theoretical results and external tools (e.g. sophisticated SAT solvers) that make their results not fully trustable and very difficult to check. To overcome this problem, a language for termination certificates, called CPF, has been developed. Deducteam develops a formally certified tool, RAINBOW, based on the Coq library COLOR, that is able to automatically verify the correctness of some of these termination certificates.

<p style="text-align:center">GALLIUM Project-Team</p>

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 OCaml 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 dereferences, 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 OCaml 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 [46] and enforcement of data confidentiality through type-based inference of information flow and noninterference properties [49].

4.3. Processing of complex structured data

Like most functional languages, OCaml 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. Therefore, OCaml 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 OCaml 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 OCaml 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 family 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.

GAMBLE Project-Team

4. Application Domains

4.1. Applications of computational geometry

Many domains of science can benefit from the results developed by GAMBLE. Curves and surfaces are ubiquitous in all sciences to understand and interpret raw data as well as experimental results. Still, the non-linear problems we address are rather basic and fundamental, and it is often difficult to predict the impact of solutions in that area. The short-term industrial impact is likely to be small because, on basic problems, industries have used ad hoc solutions for decades and have thus got used to it. The example of our work on quadric intersection is typical: even though we were fully convinced that intersecting 3D quadrics is such an elementary/fundamental problem that it ought to be useful, we were the first to be astonished by the scope of the applications of our software [0] (which was the first and still is the only one –to our knowledge– to compute robustly and efficiently the intersection of 3D quadrics) which has been used by researchers in, for instance, photochemistry, computer vision, statistics, and mathematics. Our work on certified drawing of plane (algebraic) curves falls in the same category. It seems obvious that it is widely useful to be able to draw curves correctly (recall also that part of the problem is to determine where to look in the plane) but it is quite hard to come up with specific examples of fields where this is relevant. A contrario, we know that certified meshing is critical in mechanical-design applications in robotics, which is a non-obvious application field. There, the singularities of a manipulator often have degrees higher than 10 and meshing the singular locus in a certified way is currently out of reach. As a result, researchers in robotics can only build physical prototypes for validating, or not, the approximate solutions given by non-certified numerical algorithms.

The fact that several of our pieces of software for computing non-Euclidean triangulations have already been requested by users long before they become public is a good sign for their wide future impact once in CGAL. This will not come as a surprise, since most of the questions that we have been studying followed from discussions with researchers outside computer science and pure mathematics. Such researchers are either users of our algorithms and software, or we meet them in workshops. Let us only mention a few names here. We have already referred above to our collaboration with Rien van de Weijgaert [37], [52] (astrophysicist, Groningen, NL). Michael Schindler [48] (theoretical physicist, ENSPCI, CNRS, France) is using our prototype software for 3D periodic weighted triangulations. Stephen Hyde and Vanessa Robins (applied mathematics and physics at Australian National University) have recently signed a software license agreement with INRIA that allows their group to use our prototype for 3D periodic meshing. Olivier Faugeras (neuromathematics, Inria Sophia Antipolis) had come to us and mentioned his needs for good meshes of the Bolza surface [27] before we started to study them. Such contacts are very important both to get feedback about our research and to help us choose problems that are relevant for applications. These problems are at the same time challenging from the mathematical and algorithmic points of view. Note that our research and our software are generic, i.e., we are studying fundamental geometric questions, which do not depend on any specific application. This recipe has made the sucess of the CGAL library.

Probabilistic models for geometric data are widely used to model various situations ranging from cell phone distribution to quantum mechanics. The impact of our work on probabilistic distributions is twofold. On the one hand, our studies of properties of geometric objects built on such distributions will yield a better understanding of the above phenomena and has potential impact in many scientific domains. On the other hand, our work on simulations of probabilistic distributions will be used by other teams, more maths oriented, to study these distributions.

---

[0] QI: http://vegas.loria.fr/qi/.

GRACE Project-Team (section vide)

HYCOMES Project-Team (section vide)

KAIROS Team

4. Application Domains

4.1. Model-Based Embedded System Design

Kairos general area of activity is embedded systems, as found in transportation (automotive, avionics), connected objects (smartphones,...) satellites, smart AE... More specifically we are focused on the design methods and tools to model and deploy such systems, where software is tightly coupled to its physical environment and its heterogenous running platform. Inversely we are not focused in designing new nature of cyber-physical models, but rather to study tools that help collaborative design between engineers of these distinct fields.

LFANT Project-Team (section vide)

MARELLE Project-Team (section vide)

MEXICO Project-Team

4. Application Domains

4.1. Telecommunications

Participants: Stefan Haar, Serge Haddad.

Telecommunications

MExICo's research is motivated by problems of system management in several domains, such as:

- In the domain of service oriented computing, it is often necessary to insert some Web service into an existing orchestrated business process, e.g. to replace another component after failures. This requires to ensure, often actively, conformance to the interaction protocol. One therefore needs to synthesize adaptators for every component in order to steer its interaction with the surrounding processes.

- Still in the domain of telecommunications, the supervision of a network tends to move from out-of-band technology, with a fixed dedicated supervision infrastructure, to in-band supervision where the supervision process uses the supervised network itself. This new setting requires to revisit the existing supervision techniques using control and diagnosis tools.

Currently, we have no active cooperation on thes subjects.

4.2. Biological Systems

Participants: Thomas Chatain, Stefan Haar, Serge Haddad, Stefan Schwoon.

We have begun in 2014 to examine concurrency issues in systems biology, and are currently enlarging the scope of our research's applications in this direction. To see the context, note that in recent years, a considerable shift of biologists' interest can be observed, from the mapping of static genotypes to gene expression, i.e. the processes in which genetic information is used in producing functional products. These processes are far from being uniquely determined by the gene itself, or even jointly with static properties of the environment; rather, regulation occurs throughout the expression processes, with specific mechanisms increasing or decreasing the production of various products, and thus modulating the outcome. These regulations are central in understanding cell fate (how does the cell differenciate ? Do mutations occur ? etc), and progress there hinges on our capacity to analyse, predict, monitor and control complex and variegated processes. We have applied Petri net unfolding techniques for the efficient computation of attractors in a regulatory network; that is, to identify strongly connected reachability components that correspond to stable evolutions, e.g. of a cell that differentiates into a specific functionality (or mutation). This constitutes the starting point of a broader research with Petri net unfolding techniques in regulation. In fact, ,he use of ordinary Petri nets for capturing regulatory network (RN) dynamics overcomes the limitations of traditional RN models : those impose e.g. Monotonicity properties in the influence that one factor had upon another, i.e. always increasing or always decreasing, and were thus unable to cover all actual behaviours (see [75]). Rather, we follow the more refined model of boolean networks of automata, where the local states of the different factors jointly detemine which state transitions are possible. For these connectors, ordinary PNs constitute a first approximation, improving greatly over the literature but leaving room for improvement in terms of introducing more refined logical connectors. Future work thus involves transcending this class of PN models. Via unfoldings, one has access – provided efficient techniques are available – to all behaviours of the model, rather than over-or under-approximations as previously. This opens the way to efficiently searching in particular for determinants of the cell fate : which attractors are reachable from a given stage, and what are the factors that decide in favor of one or the other attractor, etc. Our current research focusses on *cellular reprogramming*.

4.3. Autonomous Vehicles

The validation of safety properties is a crucial concern for the design of computer guided systems, in particular for automated transport systems Our approach consists in analyzing the interactions of a randomized environment (roads, cross-sections, etc.) with a vehicle controller. This requires to :

- define the relevant case studies;
- extend our tool COSMOS to handle general hybrid systems;
- conduct experimentations and analyze their results.

In [SIA2017], we have shown that this approach scales pretty well but with a controller written in C. The next step will be to combine Simulink models with Petri nets since Simulink is widely used for specifying hybrid systems in industry. In order to so, we need to define an operational semantic for Simulink and to design an elegant way for specifying the interface between nets and Simulink models. Then we will implement the solution in Cosmos.

PACAP Project-Team

4. Application Domains

4.1. Any computer usage

The PACAP team is working on the fundamental technologies for computer science: processor architecture, performance-oriented compilation and guaranteed response time for real-time. The research results may have impacts on any application domain that requires high performance execution (telecommunication, multimedia, biology, health, engineering, environment...), but also on many embedded applications that exhibit other constraints such as power consumption, code size and guaranteed response time. Our research activity implies the development of software prototypes.

PARKAS Project-Team (section vide)

PARSIFAL Project-Team

4. Application Domains

4.1. Integrating a model checker and a theorem prover

The goal of combining model checking with inductive and co-inductive theorem is appealing. The strengths of systems in these two different approaches are strikingly different. A model checker is capable of exploring a finite space automatically: such a tool can repeatedly explore all possible cases of a given computational space. On the other hand, a theorem prover might be able to prove abstract properties about a search space. For example, a model checker could attempt to discover whether or not there exists a winning strategy for, say, tic-tac-toe while an inductive theorem prover might be able to prove that if there is a winning strategy for one board then there is a winning strategy for any symmetric version of that board. Of course, the ability to combine proofs from these systems could drastically reduce the amount of state exploration and verification of proof certificates that are needed to prove the existence of winning strategies.

Our first step to providing an integration of model checking and (inductive) theorem proving was the development of a strong logic, that we call $\mathcal{G}$, which extends intuitionistic logic with notions of least and greatest fixed points. We had developed the proof theory of this logic in earlier papers [4] [57]. We have now recently converted the Bedwyr system so that it formally accepts almost all definitions and theorem statements that are accepted by the inductive theorem prover Abella. Thus, these two systems are proving theorems in the same logic and their results can now be shared.

Bedwyr's tabling mechanism has been extended so that its it can make use of previously proved lemmas. For instance, when trying to prove that some board position has a winning strategy, an available stored lemma can now be used to obtain the result if some symmetric board position is already in the table.

Heath and Miller have shown how model checking can be seen as constructing proof in (linear) logic [64]. For more about recent progress on providing checkable proof certificates for model checking, see the web site for Bedwyr http://slimmer.gforge.inria.fr/bedwyr/.

4.2. Implementing trusted proof checkers

Traditionally, theorem provers—whether interactive or automatic—are usually monolithic: if any part of a formal development was to be done in a particular theorem prover, then the whole of it would need to be done in that prover. Increasingly, however, formal systems are being developed to integrate the results returned from several, independent and high-performance, specialized provers: see, for example, the integration of Isabelle with an SMT solver [56] as well as the Why3 and ESC/Java systems.

Within the Parsifal team, we have been working on foundational aspects of this multi-prover integration problem. As we have described above, we have been developing a formal framework for defining the semantics of proof evidence. We have also been working on prototype checkers of proof evidence which are capable of executing such formal definitions. The proof definition language described in the papers [54], [53] is currently given an implementation in the λProlog programming language [74]. This initial implementation will be able to serve as a "reference" proof checker: others who are developing proof evidence definitions will be able to use this reference checker to make sure that they are getting their definitions to do what they expect.

Using λProlog as an implementation language has both good and bad points. The good points are that it is rather simple to confirm that the checker is, in fact, sound. The language also supports a rich set of abstracts which make it impossible to interfere with the code of the checker (no injection attacks are possible). On the negative side, the performance of our λProlog interpreters is lower than that of specially written checkers and kernels.

4.3. Trustworthy implementations of theorem proving techniques

Instead of integrating different provers by exchanging proof evidence and relying on a backend proof-checker, another approach to integration consists in re-implementing the theorem proving techniques as proof-search strategies, on an architecture that guarantees correctness.

Inference systems in general, and focused sequent calculi in particular, can serve as the basis of such an architecture, providing primitives for the exploration of the search space. These form a trusted *Application Programming Interface* that can be used to program and experiment various proof-search heuristics without worrying about correctness. No proof-checking is needed if one trusts the implementation of the API.

This approach has led to the development of the Psyche engine, and to its latest branch CDSAT.

Three major research directions are currently being explored, based on the above:

- The first one is about formulating automated reasoning techniques in terms of inference systems, so that they fit the approach described above. While this is rather standard for technique used in first-order Automated Theorem Provers (ATP), such as resolution, superposition, etc, this is much less standard in SMT-solving, the branch of automated reasoning that can natively handle reasoning in a combination of mathematical theories: the traditional techniques developed there usually organise the collaborations between different reasoning black boxes, whose opaque mechanisms less clearly connect to proof-theoretical inference systems. We are therefore investigating new foundations for reasoning in combinations of theories, expressed as fine-grained inference systems, and developed the *Conflict-Driven Satisfiability framework* for these foundations [19].

- The second one is about understanding how to deal with quantifiers in presence of one or more theories: On the one hand, traditional techniques for quantified problems, such as *unification* [41] or *quantifier elimination* are usually designed for either the empty theory or very specific theories. On the other hand, the industrial techniques for combining theories (Nelson-Oppen, Shostak, MCSAT [78], [82], [86], [66]) are designed for quantifier-free problems, and quantifiers there are dealt with incomplete *clause instantiation* methods or *trigger*-based techniques [55]. We are working on making the two approaches compatible.

- The above architecture's modular approach raises the question of how its different modules can safely cooperate (in terms of guaranteed correctness), while some of them are trusted and others are not. The issue is particularly acute if some of the techniques are run concurrently and exchange data at unpredictable times. For this we explore new solutions based on Milner's *LCF* [77]. In [60], we argued that our solutions in particular provide a way to fulfil the "Strategy Challenge for SMT-solving" set by De Moura and Passmore [87].

PESTO Project-Team

4. Application Domains

4.1. Formal methods for cryptographic protocols

Security protocols, such as TLS, Kerberos or ssh, are the main tool for securing our communications. The aim of our work is to propose models that are expressive enough to formally represent protocol executions in the presence of an adversary, formal definitions of the security properties to be satisfied by these protocols, and design automated tools able to analyse them and possibly exhibit design flaws.

4.2. Automated reasoning

Many techniques for symbolic verification of security are rooted in automated reasoning. A typical example is equational reasoning used to model the algebraic properties of a cryptographic primitive. Our work therefore aims to improve and adapt existing techniques or propose new ones when needed for reasoning about security.

4.3. Electronic voting

Electronic elections have in the last years been used in several countries for politically binding elections. The use in professional elections is even more widespread. The aim of our work is to increase our understanding of the security properties needed for secure elections, propose techniques for analysing e-voting protocols, design of state-of-the-art voting protocols, but also to highlight the limitations of e-voting solutions.

4.4. Privacy in social networks

Treatment of information released by users on social networks can violate a user's privacy. The goal of our work is to allow one a controlled information release while guaranteeing a user's privacy.

PI.R2 Project-Team (section vide)

POLSYS Project-Team (section vide)

<p style="text-align:center">**PRIVATICS Project-Team**</p>

3. Application Domains

3.1. Privacy in smart environments

Privacy in smart environments. One illustrative example is our latest work on privacy-preserving smart-metering [2]. Several countries throughout the world are planning to deploy smart meters in house-holds in the very near future. Traditional electrical meters only measure total consumption on a given period of time (i.e., one month or one year). As such, they do not provide accurate information of when the energy was consumed. Smart meters, instead, monitor and report consumption in intervals of few minutes. They allow the utility provider to monitor, almost in real-time, consumption and possibly adjust generation and prices according to the demand. Billing customers by how much is consumed and at what time of day will probably change consumption habits to help matching energy consumption with production. In the longer term, with the advent of smart appliances, it is expected that the smart grid will remotely control selected appliances to reduce demand. Although smart metering might help improving energy management, it creates many new privacy problems. Smart-meters provide very accurate consumption data to electricity providers. As the interval of data collected by smart meters decreases, the ability to disaggregate low-resolution data increases. Analysing high-resolution consumption data, Non-intrusive Appliance Load Monitoring (NALM) can be used to identify a remarkable number of electric appliances (e.g., water heaters, well pumps, furnace blowers, refrigerators, and air conditioners) employing exhaustive appliance signature libraries. We developed DREAM, DiffeRentially privatE smArt Metering, a scheme that is private under the differential privacy model and therefore provides strong and provable guarantees. With our scheme, an (electricity) supplier can periodically collect data from smart-meters and derive aggregated statistics while learning only limited information about the activities of individual households. For example, a supplier cannot tell from a user's trace when he watched TV or turned on heating.

3.2. Big Data and Privacy

We believe that another important problem will be related to privacy issues in big data. Public datasets are used in a variety of applications spanning from genome and web usage analysis to location-based and recommendation systems. Publishing such datasets is important since they can help us analyzing and understanding interesting patterns. For example, mobility trajectories have become widely collected in recent years and have opened the possibility to improve our understanding of large-scale social networks by investigating how people exchange information, interact, and develop social interactions. With billion of handsets in use worldwide, the quantity of mobility data is gigantic. When aggregated, they can help understand complex processes, such as the spread of viruses, and build better transportation systems. While the benefits provided by these datasets are indisputable, they unfortunately pose a considerable threat to individual privacy. In fact, mobility trajectories might be used by a malicious attacker to discover potential sensitive information about a user, such as his habits, religion or relationships. Because privacy is so important to people, companies and researchers are reluctant to publish datasets by fear of being held responsible for potential privacy breaches. As a result, only very few of them are actually released and available. This limits our ability to analyze such data to derive information that could benefit the general public. It is now an urgent need to develop Privacy-Preserving Data Analytics (PPDA) systems that collect and transform raw data into a version that is immunized against privacy attacks but that still preserves useful information for data analysis. This is one of the objectives of Privatics. There exists two classes of PPDA according to whether the entity that is collecting and anonymizing the data is trusted or not. In the trusted model, that we refer to as Privacy-Preserving Data Publishing (PPDP), individuals trust the publisher to which they disclose their data. In the untrusted model, that we refer to as Privacy-Preserving Data Collection (PPDC), individuals do not trust the data publisher. They may add some noise to their data to protect sensitive information from the data publisher.

Privacy-Preserving Data Publishing: In the trusted model, individuals trust the data publisher and disclose all their data to it. For example, in a medical scenario, patients give their true information to hospitals to receive proper treatment. It is then the responsibility of the data publisher to protect privacy of the individuals' personal data. To prevent potential data leakage, datasets must be sanitized before possible release. Several proposals have been recently proposed to release private data under the Differential Privacy model [25, 56, 26, 57, 50]. However most of these schemes release a "snapshot" of the datasets at a given period of time. This release often consists of histograms. They can, for example, show the distributions of some pathologies (such as cancer, flu, HIV, hepatitis, etc.) in a given population. For many analytics applications, "snapshots" of data are not enough, and sequential data are required. Furthermore, current work focusses on rather simple data structures, such as numerical data. Release of more complex data, such as graphs, are often also very useful. For example, recommendation systems need the sequences of visited websites or bought items. They also need to analyse people connection graphs to identify the best products to recommend. Network trace analytics also rely on sequences of events to detect anomalies or intrusions. Similarly, traffic analytics applications typically need sequences of visited places of each user. In fact, it is often essential for these applications to know that user A moved from position 1 to position 2, or at least to learn the probability of a move from position 1 to position 2. Histograms would typically represent the number of users in position 1 and position 2, but would not provide the number of users that moved from position 1 to position 2. Due to the inherent sequentiality and high-dimensionality of sequential data, one major challenge of applying current data sanitization solutions on sequential data comes from the uniqueness of sequences (e.g., very few sequences are identical). This fact makes existing techniques result in poor utility. Schemes to privately release data with complex data structures, such as sequential, relational and graph data, are required. This is one the goals of Privatics. In our current work, we address this challenge by employing a variable-length n-gram model, which extracts the essential information of a sequential database in terms of a set of variable-length n - grams [15]. We then intend to extend this approach to more complex data structures.

Privacy-Preserving Data Collection: In the untrusted model, individuals do not trust their data publisher. For example, websites commonly use third party web analytics services, such as Google Analytics to obtain aggregate traffic statistics such as most visited pages, visitors' countries, etc. Similarly, other applications, such as smart metering or targeted advertising applications, are also tracking users in order to derive aggregated information about a particular class of users. Unfortunately, to obtain this aggregate information, services need to track users, resulting in a violation of user privacy. One of our goals is to develop Privacy-Preserving Data Collection solutions. We propose to study whether it is possible to provide efficient collection/aggregation solutions without tracking users, i.e. without getting or learning individual contributions.

<div align="center">**PROSECCO Project-Team**</div>

4. Application Domains

4.1. Cryptographic Protocol Libraries

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 and SSH, 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 serve pages over HTTPS, authenticate users with a single sign-on protocol such as OAuth, encrypt user files on the server-side using XML encryption, and deploy client-side cryptographic mechanisms using a JavaScript cryptographic library. The security of these applications depends on the public key infrastructure (X.509 certificates), web browsers' implementation of HTTPS and the same origin policy (SOP), the semantics of JavaScript, HTML5, and their various associated security standards, as well as the correctness of the specific web application code of interest. We build analysis tools to find bugs in all these artifacts and verification tools that can analyze commercial web applications and evaluate their security against sophisticated web-based attacks.

<p style="text-align:center; color:red;">**SECRET Project-Team**</p>

4. Application Domains

4.1. Cryptographic primitives

Our major application domain is the design of cryptographic primitives, especially for platforms with restricting implementation requirements. For instance, we aim at recommending (or designing) low-cost (or extremely fast) encryption schemes, or primitives which remain secure against quantum computers.

4.2. Code Reconstruction

To evaluate the quality of a cryptographic algorithm, it is usually assumed that its specifications are public, as, in accordance with Kerckhoffs principle, it would be dangerous to rely, even partially, on the fact that the adversary does not know those specifications. However, this fundamental rule does not mean that the specifications are known to the attacker. In practice, before mounting a cryptanalysis, it is necessary to strip off the data. This reverse-engineering process is often subtle, even when the data formatting is not concealed on purpose. A typical case is interception: some raw data, not necessarily encrypted, is observed out of a noisy channel. To access the information, the whole communication system has first to be disassembled and every constituent reconstructed. A transmission system actually corresponds to a succession of elements (symbol mapping, scrambler, channel encoder, interleaver...), and there exist many possibilities for each of them. In addition to the "preliminary to cryptanalysis" aspect, there are other links between those problems and cryptology. They share some scientific tools (algorithmics, discrete mathematics, probability...), but beyond that, there are some very strong similarities in the techniques.

<p style="text-align:center; color:red;">**SPADES Project-Team**</p>

4. Application Domains

4.1. Industrial Applications

Our applications are in the embedded system area, typically: transportation, energy production, robotics, telecommunications, the Internet of things (IoT), systems on chip (SoC). In some areas, safety is critical, and motivates the investment in formal methods and techniques for design. But even in less critical contexts, like telecommunications and multimedia, these techniques can be beneficial in improving the efficiency and the quality of designs, as well as the cost of the programming and the validation processes.

Industrial acceptance of formal techniques, as well as their deployment, goes necessarily through their usability by specialists of the application domain, rather than of the formal techniques themselves. Hence, we are looking to propose domain-specific (but generic) realistic models, validated through experience (*e.g.*, control tasks systems), based on formal techniques with a high degree of automation (*e.g.*, synchronous models), and tailored for concrete functionalities (*e.g.*, code generation).

4.2. Industrial Design Tools

The commercially available design tools (such as UML with real-time extensions, MATLAB/ SIMULINK/ dSPACE[0]) and execution platforms (OS such as VxWORKS, QNX, real-time versions of LINUX ...) start now to provide, besides their core functionalities, design or verification methods. Some of them, founded on models of reactive systems, come close to tools with a formal basis, such as for example STATEMATE by iLogix.

Regarding the synchronous approach, commercial tools are available: SCADE[0] (based on LUSTRE), CONTROLBUILD and RT-BUILDER (based on SIGNAL) from GEENSYS[0] (part of DASSAULTSYSTEMES), specialized environments like CELLCONTROL for industrial automatism (by the INRIA spin-off ATHYS– now part of DASSAULTSYSTEMES). One can observe that behind the variety of actors, there is a real consistency of the synchronous technology, which makes sure that the results of our work related to the synchronous approach are not restricted to some language due to compatibility issues.

4.3. Current Industrial Cooperations

Regarding applications and case studies with industrial end-users of our techniques, we cooperate with Thales on schedulability analysis for evolving or underspecified real-time embedded systems, with Orange Labs on software architecture for cloud services and with Daimler on reduction of nondeterminism and analysis of deadline miss models for the design of automotive systems.

---

[0] http://www.dspaceinc.com
[0] http://www.esterel-technologies.com
[0] http://www.geensoft.com

SPECFUN Project-Team (section vide)

SUMO Project-Team

4. Application Domains

4.1. Smart transportation systems

The smart-city trend aims at optimizing all functions of future cities with the help of digital technologies. We focus on the segment of urban trains, which will evolve from static and scheduled offers to reactive and eventually on-demand transportation offers. We address two challenges in this field. The first one concerns the optimal design of robust subway lines. The idea is to be able to evaluate, at design time, the performance of time tables and of different regulations policies. In particular, we focus on robustness issues: how can small perturbations and incidents be accommodated by the system, how fast will return to normality occur, when does the system become unstable? The second challenge concerns the design of new robust regulation strategies to optimize delays, recovery times, and energy consumption at the scale of a full subway line. These problems involve large-scale discrete-event systems, with temporal and stochastic features, and translate into robustness assessment, stability analysis and joint numerical/combinatorial optimization problems on the trajectories of these systems.

4.2. Management of telecommunication networks and of data centers

Telecommunication-network management is a rich provider of research topics for the team, and some members of SUMO have a long background of contacts and transfer with industry in this domain. Networks are typical examples of large distributed dynamic systems, and their management raises numerous problems ranging from diagnosis (or root-cause analysis), to optimization, reconfiguration, provisioning, planning, verification, etc. They also bring new challenges to the community, for example on the modeling side: building or learning a network model is a complex task, specifically because these models should reflect features like the layering, the multi-resolution view of components, the description of both functions, protocols and configuration, and they should also reflect dynamically-changing architectures. Besides modeling, management algorithms are also challenged by features like the size of systems, the need to work on abstractions, on partial models, on open systems, etc. The networking technology is now evolving toward software-defined networks, virtualized-network functions, multi-tenant systems, etc., which reinforces the need for more automation in the management of such systems.

Data centers are another example of large-scale modular dynamic and reconfigurable systems: they are composed of thousands of servers, on which virtual machines are activated, migrated, resized, etc. Their management covers issues like troubleshooting, reconfiguration, optimal control, in a setting where failures are frequent and mitigated by the performance of the management plane. We have a solid background in the coordination of the various autonomic managers that supervise the different functions/layers of such systems (hardware, middleware, web services, ...) Virtualization technologies now reach the domain of networking, and telecommunication operators/vendors evolve towards providers of distributed open clouds. This convergence of IT and networking strongly calls for new management paradigms, which is an opportunity for the team.

This application domain will be revived in the team by a collaboration with Orange Labs (1 Cifre PhD in the common lab Orange/Inria) and a collaboration with Nokia Bell Labs (1 Cifre PhD, and participation to the joint research team "Softwarization of Everything" of the common lab Nokia Bell Labs/Inria).

4.3. Collaborative workflows

A current trend is to involve end-users in collection and analysis of data. Exemples of this trend are contributive science, crisis-management systems, and crowds. All these applications are data-centric and user-driven. They often are distributed and involve complex, and sometimes dynamic workflows. In many cases, there are

strong interactions between data and control flows: indeed, decisons taken regarding the next tasks to be launched highly depend on collected data. For instance, in an epidemic-surveillance system, the aggregation of various reported disease cases may trigger alerts. Another example is crowds where user skills are used to complete tasks that are better performed by humans than computers. In return, this requires addressing imprecise and sometimes unreliable answers. We address several issues related to complex workflows and data. We study declarative and dynamic models that can handle workflows, data, uncertainty, and competence management.

Once these models are mature enough, we plan to experiment them on real use cases from contributive science, health-management systems, and crowd platforms using prototypes. We also plan to define abstaction schemes allowing formal reasonning on these systems.

4.4. Systems Biology

Systems Biology is a recent topic in SUMO. In systems biology, many continuous variables interact together. Biological systems are thus good representatives for large complex quantitative systems, for which we are developing analysis and management methods. For instance, the biological pathway of apoptosis explains how many molecules interact inside a cell, triggered by some outside signal (drug, etc.), eventually leading to the death of the cell by apoptosis. While intrinsically quantitative in nature and in problems, data are usually noisy and problems need not be answered with ultimate precision. It thus seems reasonable to resort to approximations in order to handle the state-space explosion resulting from the high dimensionality of biological systems.

We are developing models and abstraction tools for systems biology. Studying these models suggests new reduction methods, such as considering populations instead of explicitly representing every single element into play (be it cells, molecules, etc): we thus develop algorithm handling population symbolically, either in a continuous (distributions) or a discrete (parametric) way. An intermediate goal is to speed-up analysis of such systems using abstractions, and a long term goal is to develop top-down model-checking methods that can be run on these abstractions.

TAMIS Team

4. Application Domains

4.1. System analysis

The work performed in Axes 1 and 2 and the methods developed there are applicable to the domain of system analysis, both wrt. program analysis and hardware analysis.

4.2. Cybersecurity

The work done in the 3 axes above aims at improving cybersecurity, be it via vulnerability analyses, malware analyses and the development of safer networking mechanisms.

4.3. Safe Internet

The work done in Axis 3 above very directly contributes to the goal of a safer Internet.

<p style="text-align:center">**TEA Project-Team**</p>

4. Application Domains

4.1. Automotive and Avionics

From our continuous collaboration with major academic and industrial partners through projects TOPCASED, OPENEMBEDD, SPACIFY, CESAR, OPEES, P and CORAIL, our experience has primarily focused on the aerospace domain. The topics of time and architecture of team TEA extend to both avionics and automotive. Yet, the research focus on time in team TEA is central in any aspect of, cyber-physical, embedded system design in factory automation, automotive, music synthesis, signal processing, software radio, circuit and system on a chip design; many application domains which, should more collaborators join the team, would definitely be worth investigating.

Multi-scale, multi-aspect time modeling, analysis and software synthesis will greatly contribute to architecture modeling in these domains, with applications to optimized (distributed, parallel, multi-core) code generation for avionics (project Corail with Thales avionics, section 8) as well as modeling standards, real-time simulation and virtual integration in automotive (project with Toyota ITC, section 8).

Together with the importance of open-source software, one of these projects, the FUI Project P (section 8), demonstrated that a centralized model for system design could not just be a domain-specific programming language, such as discrete Simulink data-flows or a synchronous language. Synchronous languages implement a fixed model of time using logical clocks that are abstraction of time as sensed by software. They correspond to a fixed viewpoint in system design, and in a fixed hardware location in the system, which is not adequate to our purpose and must be extended.

In project P, we first tried to define a centralized model for importing discrete-continuous models onto a simplified implementation of SIMULINK: P models. Certified code generators would then be developed from that format. Because this does not encompass all aspects being translated to P, the P meta-model is now being extended to architecture description concepts (of the AADL) in order to become better suited for the purpose of system design. Another example is the development of System modeler on top of SCADE, which uses the more model-engineering flavored formalism SysML to try to unambiguously represent architectures around SCADE modules.

An abstract specification formalism, capable of representing time, timing relations, with which heterogeneous models can be abstracted, from which programs can be synthesized, naturally appears better suited for the purpose of virtual prototyping. RT-Builder, based on Signal like Polychrony and developed by TNI, was industrially proven and deployed for that purpose at Peugeot. It served to develop the virtual platform simulating all on-board electronics of PSA cars. This 'hardware in the loop" simulator was used to test equipments supplied by other manufacturers with respect to virtual cars. In the advent of the related automotive standard, RT-Builder then became AUTOSAR-Builder.

4.2. Factory Automation

In collaboration with Mitsubishi R&D, we explore another application domain where time and domain heterogeneity are prime concerns: factory automation. In factory automation alone, a system is conventionally built from generic computing modules: PLCs (Programmable Logic Controllers), connected to the environment with actuators and detectors, and linked to a distributed network. Each individual, physically distributed, PLC module must be timely programmed to perform individually coherent actions and fulfill the global physical, chemical, safety, power efficiency, performance and latency requirements of the whole production chain. Factory chains are subject to global and heterogeneous (physical, electronic, functional) requirements whose enforcement must be orchestrated for all individual components.

Model-based analysis in factory automation emerges from different scientific domains and focus on different CPS abstractions that interact in subtle ways: logic of PLC programs, real-time electromechanical processing, physical and chemical environments. This yields domain communication problems that render individual domain analysis useless. For instance, if one domain analysis (e.g. software) modifies a system model in a way that violates assumptions made by another domain (e.g. chemistry) then the detection of its violation may well be impossible to explain to either of the software and chemistry experts. As a consequence, cross-domain analysis issues are discovered very late during system integration and lead to costly fixes. This is particularly prevalent in multi-tier industries, such as avionic, automotive, factories, where systems are prominently integrated from independently-developed parts.

<p style="text-align:center">**TOCCATA Project-Team**</p>

4. Application Domains

4.1. Domain

The application domains we target involve safety-critical software, that is where a high-level guarantee of soundness of functional execution of the software is wanted. Currently our industrial collaborations mainly belong to the domain of transportation, including aeronautics, railroad, space flight, automotive.

Verification of C programs, Alt-Ergo at Airbus Transportation is the domain considered in the context of the ANR U3CAT project, led by CEA, in partnership with Airbus France, Dassault Aviation, Sagem Défense et Sécurité. It included proof of C programs via Frama-C/Jessie/Why, proof of floating-point programs [114], the use of the Alt-Ergo prover via CAVEAT tool (CEA) or Frama-C/WP. Within this context, we contributed to a qualification process of Alt-Ergo with Airbus industry: the technical documents (functional specifications and benchmark suite) have been accepted by Airbus, and these documents were submitted by Airbus to the certification authorities (DO-178B standard) in 2012. This action is continued in the new project Soprano.

Certified compilation, certified static analyzers Aeronautics is the main target of the Verasco project, led by Verimag, on the development of certified static analyzers, in partnership with Airbus. This is a follow-up of the transfer of the CompCert certified compiler (Inria team Gallium) to which we contributed to the support of floating-point computations [64].

Transfer to the community of Ada development The former FUI project Hi-Lite, led by Adacore company, introduced the use of Why3 and Alt-Ergo as back-end to SPARK2014, an environment for verification of Ada programs. This is applied to the domain of aerospace (Thales, EADS Astrium). At the very beginning of that project, Alt-Ergo was added in the Spark Pro toolset (predecessor of SPARK2014), developed by Altran-Praxis: Alt-Ergo can be used by customers as an alternate prover for automatically proving verification conditions. Its usage is described in the new edition of the Spark book [0] (Chapter "Advanced proof tools"). This action is continued in the new joint laboratory ProofInUse. A recent paper [72] provides an extensive list of applications of SPARK, a major one being the British air control management *iFacts*.

Transfer to the community of Atelier B In the current ANR project BWare, we investigate the use of Why3 and Alt-Ergo as an alternative back-end for checking proof obligations generated by *Atelier B*, whose main applications are railroad-related software,

a collaboration with Mitsubishi Electric R&D Centre Europe (Rennes) (joint publication [119]) and ClearSy (Aix-en-Provence).

SMT-based Model-Checking: Cubicle S. Conchon (with A. Mebsout and F. Zaidi from VALS team at LRI) has a long-term collaboration with S. Krstic and A. Goel (Intel Strategic Cad Labs in Hillsboro, OR, USA) that aims in the development of the SMT-based model checker Cubicle (http://cubicle.lri.fr/) based on Alt-Ergo [116][5]. It is particularly targeted to the verification of concurrent programs and protocols.

---

[0]http://www.altran-praxis.com/book/

VERIDIS Project-Team

4. Application Domains

4.1. Application Domains

Distributed algorithms and protocols are found at all levels of computing infrastructure, from many-core processors and systems-on-chip to wide-area networks. We are particularly interested in the verification of algorithms that are developed for supporting novel computing paradigms, including ad-hoc networks that underly mobile and low-power computing or overlay networks, peer-to-peer networking that provide services for telecommunication or cloud computing services. Computing infrastructure must be highly available and is ideally invisible to the end user, therefore correctness is crucial. One should note that standard problems of distributed computing such as consensus, group membership or leader election have to be reformulated for the dynamic context of these modern systems. We are not ourselves experts in the design of distributed algorithms, but we work together with domain experts on designing formal models of these protocols, and on verifying their properties. These cooperations help us focus on concrete algorithms and ensure that our work is relevant to the distributed algorithm community.

Formal verification techniques can contribute to certifying the correctness of systems. In particular, they help assert under which assumptions an algorithm or system functions as required. For example, the highest levels of the Common Criteria for Information Technology Security Evaluation encourage the use of formal methods. While initially the requirements of certified development have mostly been restricted to safety-critical systems, the cost of unavailable services due to malfunctioning system components and software provides wider incentives for verification. For example, we have been working on modeling and verifying medical devices that require closed-loop models of both the system and its environment.

<center>ACUMES Project-Team</center>

4. Application Domains

4.1. Active flow control for vehicles

The reduction of CO_2 emissions represents a great challenge for the automotive and aeronautic industries, which committed respectively a decrease of 20% for 2020 and 75% for 2050. This goal will not be reachable, unless a significant improvement of the aerodynamic performance of cars and aircrafts is achieved (e.g. aerodynamic resistance represents 70% of energy losses for cars above 90 km/h). Since vehicle design cannot be significantly modified, due to marketing or structural reasons, active flow control technologies are one of the most promising approaches to improve aerodynamic performance. This consists in introducing micro-devices, like pulsating jets or vibrating membranes, that can modify vortices generated by vehicles. Thanks to flow non-linearities, a small energy expense for actuation can significantly reduce energy losses. The efficiency of this approach has been demonstrated, experimentally as well as numerically, for simple configurations [147]. However, the lack of efficient and flexible numerical models, that allow to simulate and optimize a large number of such devices on realistic configurations, is still a bottleneck for the emergence of this technology in an industrial context. In particular, the prediction of actuated flows requires the use of advanced turbulence closures, like Detached Eddy Simulation or Large Eddy Simulation [96]. They are intrinsically three-dimensional and unsteady, yielding a huge computational effort for each analysis, which makes their use tedious for optimization purpose. In this context, we intend to contribute to the following research axes:

- *Sensitivity analysis for actuated flows.* Adjoint-based (reverse) approaches, classically employed in design optimization procedure to compute functional gradients, are not well suited to this context. Therefore, we propose to explore the alternative (direct) formulation, which is not so much used, in the perspective of a better characterization of actuated flows and optimization of control devices.

- *Hierarchical optimization of control devices.* The optimization of dozen of actuators, in terms of locations, frequencies, amplitudes, will be practically tractable only if a hierarchical approach is adopted, which mixes fine (DES) and coarse (URANS) simulations, and possibly experiments. We intend to develop such an optimization strategy on the basis of Gaussian Process models (*multi-fidelity kriging*).

4.2. Vehicular and pedestrian traffic flows

Intelligent Transportation Systems (ITS) is nowadays a booming sector, where the contribution of mathematical modeling and optimization is widely recognized. In this perspective, traffic flow models are a commonly cited example of "complex systems", in which individual behavior and self-organization phenomena must be taken into account to obtain a realistic description of the observed macroscopic dynamics [106]. Further improvements require more advanced models, keeping into better account interactions at the microscopic scale, and adapted control techniques, see [56] and references therein. In particular, we will focus on the following aspects:

- *Junction models.* We are interested in designing a general junction model both satisfying basic analytical properties guaranteeing well-posedness and being realistic for traffic applications. In particular, the model should be able to overcome severe drawbacks of existing models, such as restrictions on the number of involved roads and prescribed split ratios [69], [94], which limit their applicability to real world situations. Hamilton-Jacobi equations could be also an interesting direction of research, following the recent results obtained in [111].

- *Data assimilation.* In traffic flow modeling, the capability of correctly estimating and predicting the state of the system depends on the availability of rich and accurate data on the network. Up to now, the most classical sensors are fixed ones. They are composed of inductive loops (electrical wires) that are installed at different spatial positions of the network and that can measure the traffic flow,

the occupancy rate (i.e. the proportion of time during which a vehicle is detected to be over the loop) and the speed (in case of a system of two distant loops). These data are useful / essential to calibrate the phenomenological relationship between flow and density which is known in the traffic literature as the Fundamental Diagram. Nowadays, thanks to the wide development of mobile internet and geolocalization techniques and its increasing adoption by the road users, smartphones have turned into perfect mobile sensors in many domains, including in traffic flow management. They can provide the research community with a large database of individual trajectory sets that are known as Floating Car Data (FCD), see [108] for a real field experiment. Classical macroscopic models, say (hyperbolic systems of) conservation laws, are not designed to take into account this new kind of microscopic data. Other formulations, like Hamilton-Jacobi partial differential equations, are most suited and have been intensively studied in the past five years (see [62], [63]), with a stress on the (fixed) Eulerian framework. Up to our knowledge, there exist a few studies in the time-Lagrangian as well as space-Lagrangian frameworks, where data coming from mobile sensors could be easily assimilated, due to the fact that the Lagrangian coordinate (say the label of a vehicle) is fixed.

- *Control of autonomous vehicles.* Traffic flow is usually controlled via traffic lights or variable speed limits, which have fixed space locations. The deployment of autonomous vehicles opens new perspectives in traffic management, as the use of a small fraction of cars to optimize the overall traffic. In this perspective, the possibility to track vehicles trajectories either by coupled micro-macro models [76], [95] or via the Hamilton-Jacobi approach [62], [63] could allow to optimize the flow by controlling some specific vehicles corresponding to internal conditions.

4.3. Concurrent design for building systems

Building industry has to face more and more stringent requirements, including energy performance, structural safety and environmental impact. To this end, new materials and new technologies have emerged   [115] to help the construction firms meet these requirements. At the same time, many different teams or firms interact, most of the interaction being of non-cooperative nature. The teams involved in construction have different goals, depending on which stage they operate. Indeed, the lifetime of a building goes through three stages: construction, use and destruction. To each of these phases correspond quality criteria related in particular to:

- Safety: structural, fire, evacuation, chemical spread, etc.
- Well-being of its occupants: thermal and acoustic comfort.
- Functionality of its intended use.
- Environmental impact.

These stages and criteria form a complex system, the so-called building system, whose overall quality (in an intuitive sense) is directly impacted by many heterogeneous factors, such as the geographical location or the shape or material composition of some of its components (windows, frames, thermal convectors positions, etc.) It is obvious that the optimization process of these settings must be performed at the "zero" stage of the project design. Moreover, the optimization process has to follow a global approach, taking into account all the concurrent criteria that intervene in the design of building systems.

The application of up-to-date concurrent optimization machinery (games, Pareto Fronts) for multiphysics systems involved in the building is an original approach. With our industrial partner, who wishes routine use of new high performance components in the construction of buildings, we expect that our approach will yield breakthrough performances (with respect to the above criteria) compared to the current standards.

The research project relies on the ADT BuildingSmart (see software development section) for the implementation of industrial standard software demonstrators.

4.4. Other application fields

Besides the above mentioned axes, which constitute the project's identity, the methodological tools described in Section have a wider range of application. We currently carry on also the following research actions, in collaboration with external partners.

- **Modeling cell dynamics.** Migration and proliferation of epithelial cell sheets are the two keystone aspects of the collective cell dynamics in most biological processes such as morphogenesis, embryogenesis, cancer and wound healing. It is then of utmost importance to understand their underlying mechanisms.

  Semilinear reaction-diffusion equations are widely used to give a phenomenological description of the temporal and spatial changes occurring within cell populations that undergo scattering (moving), spreading (expanding cell surface) and proliferation. We have followed the same methodology and contributed to assess the validity of such approaches in different settings (cell sheets [103], dorsal closure [44], actin organization [43]). However, epithelial cell-sheet movement is complex enough to undermine most of the mathematical approaches based on *locality*, that is mainly traveling wavefront-like partial differential equations. In [89] it is shown that Madin-Darby Canine Kidney (MDCK) cells extend cryptic lamellipodia to drive the migration, several rows behind the wound edge. In [130] MDCK monolayers are shown to exhibit similar non-local behavior (long range velocity fields, very active border-localized leader cells).

  Our aim is to start from a mesoscopic description of cell interaction: considering cells as independent anonymous agents, we plan to investigate the use of mathematical techniques adapted from the mean-field game theory. Otherwise, looking at them as interacting particles, we will use a multi-agent approach (at least for the actin dynamics). We intend also to consider approaches stemming from compartment-based simulation in the spirit of those developed in [86], [91], [93].

- **Modeling cardio-stents.**

  Atherosclerosis or arterial calcification is a major vascular disease, caused by fatty deposits on the inner walls of arteries. Angioplasty techniques propose several solutions to remedy this pathology. We are interested in those which consist in introducing a metallic stent, to crush the lipid plaques, and ensure permanent enlargement of the damaged arterial wall. The implementation of such an element is accompanied by an immune reaction of the arterial walls, which is manifested by an accelerated proliferation of cells within the so called media, which highlights two major risks: restenosis, and thrombosis . One promising technique is to introduce a "Drug Eluting Stent", which is a metallic stent coated with a polymer layer containing an antiproliferative drug to slow the proliferation process, in order to improve the functioning of the stent. Our major objective in this part is to setup and develop the mathematical modeling and computational tools that lead to the effective estimation of the Fractional Flow Reserve [128], which is a promising new technique to help the cardiologists take decisions on stent implantation.

- **Game strategies for thermoelastography.** Thermoelastography is an innovative non-invasive control technology, which has numerous advantages over other techniques, notably in medical imaging [122]. Indeed, it is well known that most pathological changes are associated with changes in tissue stiffness, while remaining isoechoic, and hence difficult to detect by ultrasound techniques. Based on elastic waves and heat flux reconstruction, thermoelastography shows no destructive or aggressive medical sequel, unlike X-ray and comparables techniques, making it a potentially prominent choice for patients.

  Physical principles of thermoelastography originally rely on dynamical structural responses of tissues, but as a first approach, we only consider static responses of linear elastic structures.
  The mathematical formulation of the thermoelasticity reconstruction is based on data completion and material identification, making it a harsh ill posed inverse problem. In previous works [104], [113], we have demonstrated that Nash game approaches are efficient to tackle ill-posedness. We intend to extend the results obtained for Laplace equations in [104], and the algorithms developed in Section 3.1.2.4 to the following problems (of increasing difficulty):
  - Simultaneous data and parameter recovery in linear elasticity, using the so-called Kohn and Vogelius functional (ongoing work, some promising results obtained).
  - Data recovery in coupled heat-thermoelasticity systems.

- Data recovery in linear thermoelasticity under stochastic heat flux, where the imposed flux is stochastic.
- Data recovery in coupled heat-thermoelasticity systems under stochastic heat flux, formulated as an incomplete information Nash game.
- Application to robust identification of cracks.

- **Constraint elimination in Quasi-Newton methods.** In single-objective differentiable optimization, Newton's method requires the specification of both gradient and Hessian. As a result, the convergence is quadratic, and Newton's method is often considered as the target reference. However, in applications to distributed systems, the functions to be minimized are usually "functionals", which depend on the optimization variables by the solution of an often complex set of PDE's, through a chain of computational procedures. Hence, the exact calculation of the full Hessian becomes a complex and costly computational endeavor.

  This has fostered the development of *quasi-Newton's methods* that mimic Newton's method but use only the gradient, the Hessian being iteratively constructed by successive approximations inside the algorithm itself. Among such methods, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is well-known and commonly employed. In this method, the Hessian is corrected at each new iteration by rank-one matrices defined from several evaluations of the gradient only. The BFGS method has "super-linear convergence".

  For constrained problems, certain authors have developed so-called *Riemannian BFGS*, e.g. [133], that have the desirable convergence property in constrained problems. However, in this approach, the constraints are assumed to be known formally, by explicit expressions.

  In collaboration with ONERA-Meudon, we are exploring the possibility of representing constraints, in successive iterations, through local approximations of the constraint surfaces, splitting the design space locally into tangent and normal sub-spaces, and eliminating the normal coordinates through a linearization, or more generally a finite expansion, and applying the BFGS method through dependencies on the coordinates in the tangent subspace only. Preliminary experiments on the difficult Rosenbrock test-case, although in low dimensions, demonstrate the feasibility of this approach. On-going research is on theorizing this method, and testing cases of higher dimensions.

- **Multi-objective optimization for nanotechnologies.** Our team takes part in a larger collaboration with CEA/LETI (Grenoble), initiated by the Inria Project-Team Nachos, and related to the Maxwell equations. Our component in this activity relates to the optimization of nanophotonic devices, in particular with respect to the control of thermal loads. We have first identified a gradation of representative test-cases of increasing complexity:
  - infrared micro-source;
  - micro-photoacoustic cell;
  - nanophotonic device.

  These cases involve from a few geometric parameters to be optimized to a functional minimization subject to a finite-element solution involving a large number of dof's. CEA disposes of such codes, but considering the computational cost of the objective functions in the complex cases, the first part of our study is focused on the construction and validation of meta-models, typically of RBF-type. Multi-objective optimization will be carried out subsequently by MGDA, and possibly Nash games.

APICS Project-Team

4. Application Domains

4.1. Introduction

Application domains are naturally linked to the problems described in Sections 3.2.1 and 3.2.2 . By and large, they split into a systems-and-circuits part and an inverse-source-and-boundary-problems part, united under a common umbrella of function-theoretic techniques as described in Section 3.3 .

4.2. Inverse magnetization problems

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Konstantinos Mavreas.

Generally speaking, inverse potential problems, similar to the one appearing in Section 4.3 , occur naturally in connection with systems governed by Maxwell's equation in the quasi-static approximation regime. In particular, they arise in magnetic reconstruction issues. A specific application is to geophysics, which led us to form the Inria Associate Team IMPINGE (Inverse Magnetization Problems IN GEosciences) together with MIT and Vanderbilt University. A recent collaboration with Cerege (CNRS, Aix-en-Provence), in the framework of the ANR-project MagLune, completes this picture, see Section 7.2.2 .

To set up the context, recall that the Earth's geomagnetic field is generated by convection of the liquid metallic core (geodynamo) and that rocks become magnetized by the ambient field as they are formed or after subsequent alteration. Their remanent magnetization provides records of past variations of the geodynamo, which is used to study important processes in Earth sciences like motion of tectonic plates and geomagnetic reversals. Rocks from Mars, the Moon, and asteroids also contain remanent magnetization which indicates the past presence of core dynamos. Magnetization in meteorites may even record fields produced by the young sun and the protoplanetary disk which may have played a key role in solar system formation.

For a long time, paleomagnetic techniques were only capable of analyzing bulk samples and compute their net magnetic moment. The development of SQUID microscopes has recently extended the spatial resolution to sub-millimeter scales, raising new physical and algorithmic challenges. The associate team IMPINGE aims at tackling them, experimenting with the SQUID microscope set up in the Paleomagnetism Laboratory of the department of Earth, Atmospheric and Planetary Sciences at MIT. Typically, pieces of rock are sanded down to a thin slab, and the magnetization has to be recovered from the field measured on a planar region at small distance from the slab.

Mathematically speaking, both inverse source problems for EEG from Section 4.3 and inverse magnetization problems described presently amount to recover the (3-D valued) quantity m (primary current density in case of the brain or magnetization in case of a thin slab of rock) from measurements of the potential:

$$\int_{\Omega} \frac{\operatorname{div} m(x')\, dx'}{|x-x'|}\,, \tag{1}$$

outside the volume Ω of the object. The difference is that the distribution m is located in a volume in the case of EEG, and on a plane in the case of rock magnetization. This results in quite different identifiability properties, see [38] and Section 5.1.1 , but the two situations share a substantial Mathematical common core.

Another timely instance of inverse magnetization problems lies with geomagnetism. Satellites orbiting around the Earth measure the magnetic field at many points, and nowadays it is a challenge to extract global information from those measurements. In collaboration with C. Gerhards from the University of Vienna, Apics has started to work on the problem of separating the magnetic field due to the magnetization of the globe's crust from the magnetic field due to convection in the liquid metallic core. The techniques involved are variants, in a spherical context, from those developed within the IMPINGE associate team for paleomagnetism, see Section 5.1.1 .

4.3. Inverse source problems in EEG

Participants: Laurent Baratchart, Juliette Leblond, Jean-Paul Marmorat, Christos Papageorgakis.

This work is conducted in collaboration with Maureen Clerc and Théo Papadopoulo from the team Athena (Inria Sophia).

Solving overdetermined Cauchy problems for the Laplace equation on a spherical layer (in 3-D) in order to extrapolate incomplete data (see Section 3.2.1) is a necessary ingredient of the team's approach to inverse source problems, in particular for applications to EEG, see [7]. Indeed, the latter involves propagating the initial conditions through several layers of different conductivities, from the boundary shell down to the center of the domain where the singularities (*i.e.* the sources) lie. Once propagated to the innermost sphere, it turns out that traces of the boundary data on 2-D cross sections coincide with analytic functions with branched singularities in the slicing plane [6], [43]. The singularities are related to the actual location of the sources, namely their moduli reach in turn a maximum when the plane contains one of the sources. Hence we are back to the 2-D framework of Section 3.3.3 , and recovering these singularities can be performed *via* best rational approximation. The goal is to produce a fast and sufficiently accurate initial guess on the number and location of the sources in order to run heavier descent algorithms on the direct problem, which are more precise but computationally costly and often fail to converge if not properly initialized. Our belief is that such a localization process can add a geometric, valuable piece of information to the standard temporal analysis of EEG signal records.

Numerical experiments obtained with our software FindSources3D give very good results on simulated data and we are now engaged in the process of handling real experimental data (see Sections 3.4.2 and 5.1), in collaboration with the team Athena at Inria Sophia Antipolis, neuroscience teams in partner-hospitals (Institut de Neurosciences des Systèmes, hospital la Timone, Aix-Marseille Univ., Marseille, http://ins.univ-amu.fr/), and the BESA company (Munich).

4.4. Identification and design of microwave devices

Participants: Laurent Baratchart, Sylvain Chevillard, Jean-Paul Marmorat, Martine Olivi, Fabien Seyfert.

This is joint work with Stéphane Bila (XLIM, Limoges).

One of the best training grounds for function-theoretic applications by the team is the identification and design of physical systems whose performance is assessed frequency-wise. This is the case of electromagnetic resonant systems which are of common use in telecommunications.

In space telecommunications (satellite transmissions), constraints specific to on-board technology lead to the use of filters with resonant cavities in the microwave range. These filters serve multiplexing purposes (before or after amplification), and consist of a sequence of cylindrical hollow bodies, magnetically coupled by irises (orthogonal double slits). The electromagnetic wave that traverses the cavities satisfies the Maxwell equations, forcing the tangent electrical field along the body of the cavity to be zero. A deeper study of the Helmholtz equation states that an essentially discrete set of wave vectors is selected. In the considered range of frequency, the electrical field in each cavity can be decomposed along two orthogonal modes, perpendicular to the axis of the cavity (other modes are far off in the frequency domain, and their influence can be neglected).

Each cavity (see Figure 1) has three screws, horizontal, vertical and midway (horizontal and vertical are two arbitrary directions, the third direction makes an angle of 45 or 135 degrees, the easy case is when all cavities show the same orientation, and when the directions of the irises are the same, as well as the input and output slits). Since screws are conductors, they behave as capacitors; besides, the electrical field on the surface has to be zero, which modifies the boundary conditions of one of the two modes (for the other mode, the electrical field is zero hence it is not influenced by the screw), the third screw acts as a coupling between the two modes. The effect of an iris is opposite to that of a screw: no condition is imposed on a hole, which results in a coupling between two horizontal (or two vertical) modes of adjacent cavities (in fact the iris is the union of two rectangles, the important parameter being their width). The design of a filter consists in finding the size of each cavity, and the width of each iris. Subsequently, the filter can be constructed and tuned by adjusting the

Figure 1. Picture of a 6-cavities dual mode filter. Each cavity (except the last one) has 3 screws to couple the modes within the cavity, so that 16 quantities must be optimized. Quantities such as the diameter and length of the cavities, or the width of the 11 slits are fixed during the design phase.

screws. Finally, the screws are glued once a satisfactory response has been obtained. In what follows, we shall consider a typical example, a filter designed by the CNES in Toulouse, with four cavities near 11 GHz.

Near the resonance frequency, a good approximation to the Helmholtz equations is given by a second order differential equation. Thus, one obtains an electrical model of the filter as a sequence of electrically-coupled resonant circuits, each circuit being modeled by two resonators, one per mode, the resonance frequency of which represents the frequency of a mode, and whose resistance accounts for electric losses (surface currents) in the cavities.

This way, the filter can be seen as a quadripole, with two ports, when plugged onto a resistor at one end and fed with some potential at the other end. One is now interested in the power which is transmitted and reflected. This leads one to define a scattering matrix S, which may be considered as the transfer function of a stable causal linear dynamical system, with two inputs and two outputs. Its diagonal terms $S_{1,1}$, $S_{2,2}$ correspond to reflections at each port, while $S_{1,2}$, $S_{2,1}$ correspond to transmission. These functions can be measured at certain frequencies (on the imaginary axis). The matrix S is approximately rational of order 4 times the number of cavities (that is 16 in the example on Figure 2), and the key step consists in expressing the components of the equivalent electrical circuit as functions of the S_{ij} (since there are no formulas expressing the lengths of the screws in terms of parameters of this electrical model). This representation is also useful to analyze the numerical simulations of the Maxwell equations, and to check the quality of a design, in particular the absence of higher resonant modes.

In fact, resonance is not studied via the electrical model, but via a low-pass equivalent circuit obtained upon linearizing near the central frequency, which is no longer conjugate symmetric (*i.e.* the underlying system may no longer have real coefficients) but whose degree is divided by 2 (8 in the example).

In short, the strategy for identification is as follows:

- measuring the scattering matrix of the filter near the optimal frequency over twice the pass band (which is 80MHz in the example).
- Solving bounded extremal problems for the transmission and the reflection (the modulus of he

response being respectively close to 0 and 1 outside the interval measurement, cf. Section 3.3.1) in order to get a models for the scattering matrix as an analytic matrix-valued function. This provides us with a scattering matrix known to be close to a rational matrix of order roughly 1/4 of the number of data points.

- Approximating this scattering matrix by a true rational transfer-function of appropriate degree (8 in this example) via the Endymion or RARL2 software (cf. Section 3.3.2.2).

- A state space realization of S, viewed as a transfer function, can then be obtained, where additional symmetry constraints coming from the reciprocity law and possibly other physical features of the device have to be imposed.

- Finally one builds a realization of the approximant and looks for a change of variables that eliminates non-physical couplings. This is obtained by using algebraic-solvers and continuation algorithms on the group of orthogonal complex matrices (symmetry forces this type of transformation).
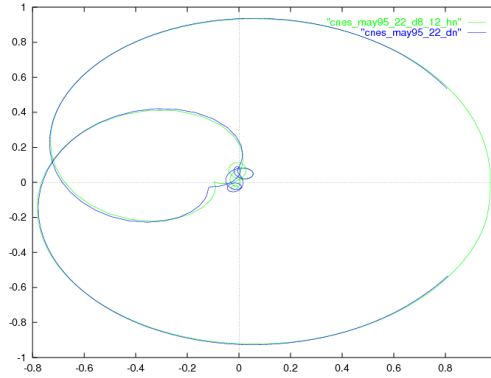
Figure 2. Nyquist Diagram. Rational approximation (degree 8) and data - S_{22}.

The final approximation is of high quality. This can be interpreted as a confirmation of the linearity assumption on the system: the relative L^2 error is less than 10^{-3}. This is illustrated by a reflection diagram (Figure 2). Non-physical couplings are less than 10^{-2}.

The above considerations are valid for a large class of filters. These developments have also been used for the design of non-symmetric filters, which are useful for the synthesis of repeating devices.

The team further investigates problems relative to the design of optimal responses for microwave devices. The resolution of a quasi-convex Zolotarev problems was proposed, in order to derive guaranteed optimal multi-band filter responses subject to modulus constraints [8]. This generalizes the classical single band design techniques based on Chebyshev polynomials and elliptic functions. The approach relies on the fact that the modulus of the scattering parameter $|S_{1,2}|$ admits a simple expression in terms of the filtering function $D = |S_{1,1}|/|S_{1,2}|$, namely

$$|S_{1,2}|^2 = \frac{1}{1 + D^2}.$$

The filtering function appears to be the ratio of two polynomials p_1/p_2, the numerator of the reflection and transmission scattering factors, that may be chosen freely. The denominator q is then obtained as the unique stable unitary polynomial solving the classical Feldtkeller spectral equation:

$$qq^* = p_1 p_1^* + p_2 p_2^*.$$

The relative simplicity of the derivation of a filter's response, under modulus constraints, owes much to the possibility of forgetting about Feldtkeller's equation and express all design constraints in terms of the filtering function. This no longer the case when considering the synthesis N-port devices for $N > 3$, like multiplexers, routers and power dividers, or when considering the synthesis of filters under matching conditions. The efficient derivation of multiplexers responses is the subject of recent investigation by Apics, using techniques based on constrained Nevanlinna-Pick interpolation (see Section 5.2).

Through contacts with CNES (Toulouse) and UPV (Bilbao), Apics got additionally involved in the design of amplifiers which, unlike filters, are active devices. A prominent issue here is stability. A twenty years back, it was not possible to simulate unstable responses, and only after building a device could one detect instability. The advent of so-called *harmonic balance* techniques, which compute steady state responses of linear elements in the frequency domain and look for a periodic state in the time domain of a network connecting these linear elements *via* static non-linearities made it possible to compute the harmonic response of a (possibly nonlinear and unstable) device [80]. This has had tremendous impact on design, and there is a growing demand for software analyzers. The team is also becoming active in this area.

In this connection, there are two types of stability involved. The first is stability of a fixed point around which the linearized transfer function accounts for small signal amplification. The second is stability of a limit cycle which is reached when the input signal is no longer small and truly nonlinear amplification is attained (*e.g.* because of saturation). Work by the team so far has been concerned with the first type of stability, and emphasis is put on defining and extracting the "unstable part" of the response, see Section 5.3 . The stability check for limit cycles is now under investigation.

<p style="text-align:center">**ASPI Team**</p>

4. Application Domains

4.1. Localisation, navigation and tracking

See 6.1 .

Among the many application domains of particle methods, or interacting Monte Carlo methods, ASPI has decided to focus on applications in localisation (or positioning), navigation and tracking [39], [32], which already covers a very broad spectrum of application domains. The objective here is to estimate the position (and also velocity, attitude, etc.) of a mobile object, from the combination of different sources of information, including

- a prior dynamical model of typical evolutions of the mobile, such as inertial estimates and prior model for inertial errors,
- measurements provided by sensors,
- and possibly a digital map providing some useful feature (terrain altitude, power attenuation, etc.) at each possible position.

In some applications, another useful source of information is provided by

- a map of constrained admissible displacements, for instance in the form of an indoor building map,

which particle methods can easily handle (map-matching). This Bayesian dynamical estimation problem is also called filtering, and its numerical implementation using particle methods, known as particle filtering, has been introduced by the target tracking community [38], [51], which has already contributed to many of the most interesting algorithmic improvements and is still very active, and has found applications in

> target tracking, integrated navigation, points and / or objects tracking in video sequences, mobile robotics, wireless communications, ubiquitous computing and ambient intelligence, sensor networks, etc.

ASPI is contributing (or has contributed recently) to several applications of particle filtering in positioning, navigation and tracking, such as geolocalisation and tracking in a wireless network, terrain–aided navigation, and data fusion for indoor localisation.

4.2. Rare event simulation

See 3.2 , 5.1 , and 5.2 .

Another application domain of particle methods, or interacting Monte Carlo methods, that ASPI has decided to focus on is the estimation of the small probability of a rare but critical event, in complex dynamical systems. This is a crucial issue in industrial areas such as

> nuclear power plants, food industry, telecommunication networks, finance and insurance industry, air traffic management, etc.

In such complex systems, analytical methods cannot be used, and naive Monte Carlo methods are clearly unefficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [45], where trajectories going towards the critical set are given offsprings, thus increasing the number of trajectories that eventually reach the critical set. This approach not only makes it possible to estimate the probability of the rare event, but also provides realizations of the random trajectory, given that it reaches the critical set, i.e. provides realizations of typical critical trajectories, an important feature that methods based on importance sampling usually miss.

ASPI is contributing (or has contributed recently) to several applications of multilevel splitting for rare event simulation, such as risk assessment in air traffic management, detection in sensor networks, and protection of digital documents.

BIPOP Project-Team

4. Application Domains

4.1. Computational neuroscience

Modeling in neuroscience makes extensive use of nonlinear dynamical systems with a huge number of interconnected elements. Our current theoretical understanding of the properties of neural systems is mainly based on numerical simulations, from single cell models to neural networks. To handle correctly the discontinuous nature of integrate-and-fire networks, specific numerical schemes have to be developed. Our current works focus on event-driven, time-stepping and voltage-stepping strategies, to simulate accurately and efficiently neuronal networks. Our activity also includes a mathematical analysis of the dynamical properties of neural systems. One of our aims is to understand neural computation and to develop it as a new type of information science [18], [19].

4.2. Electronic circuits

Whether they are integrated on a single substrate or as a set of components on a board, electronic circuits are very often a complex assembly of many basic components with non linear characteristics. The IC technologies now allow the integration of hundreds of millions of transistors switching at GHz frequencies on a die of 1cm^2. It is out of the question to simulate a complete IC with standard tools such as the SPICE simulator. We currently work on a dedicated plug-in able to simulate a whole circuit comprising various components, some modelled in a nonsmooth way [1].

4.3. Walking robots

As compared to rolling robots, the walking ones – for example hexapods – possess definite advantages whenever the ground is not flat or free: clearing obstacles is easier, holding on the ground is lighter, adaptivity is improved. However, if the working environment of the system is adapted to man, the biped technology must be preferred, to preserve good displacement abilities without modifying the environment. This explains the interest displayed by the international community in robotics toward humanoid systems, whose aim is to back man in some of his activities, professional or others. For example, a certain form of help at home to disabled persons could be done by biped robots, as they are able to move without any special adaptation of the environment.

4.4. Computer graphics animation

Computer graphics animation is dedicated to the numerical modeling and simulation of physical phenomena featuring a high visual impact. Typically, deformable objects prone to strong deformation, large displacements, complex and nonlinear or even nonsmooth behavior, are of interest for this community. We are interested in two main mechanical phenomena: on the one hand, the behavior of slender (nonlinear) structures such as rods, plates and shells; on the other hand, the effect of frictional contact between rigid or deformable bodies. In both cases the goal is to design realistic, efficient, robust, and controllable computational models. Whereas the problem of collision detection has become a mature field those recent years, simulating the collision response (in particular frictional contacts) in a realistic, robust and efficient way, still remains an important challenge. We have focussed in the past years on the simulation of heterogeneous objects such as granular or fibrous materials, both with a discrete element point of view [11], and, more recently, with a macroscopic (continuum) point of view [12]. We also pursue some study on the design of high-order models for slender structures such as rods, plates or shells. Our current activity includes the static inversion of mechanical objects, which is of great importance in the field of artistic design, for the making of movies and video games for example. Such problems typically involve geometric fitting and parameters identification issues, both resolved with the help of constrained optimization. Finally, we are interested in studying certain discrepancies (inexistence of solution) due to the combination of incompatible models such as contacting rigid bodies subject to Coulomb friction.

4.5. Multibody Systems: Modeling, Control, Waves, Simulation

Multibody systems are assemblies of rigid or flexible bodies, typically modeled with Newton-Euler or Lagrange dynamics, with bilateral and unilateral constraints, with or without tangential effects like friction. These systems are highly nonlinear and nonsmooth, and are therefore challenging for modeling aspects (impact dynamics, especially multiple –simultaneous– collisions), feedback control [10], state observation, as well as numerical analysis and simulation (software development) [2], [4], [5]. Biped robots are a particular, interesting subclass of multibody systems subject to various constraints. Granular materials are another important field, in which nonlinear waves transmissions are crucial (one celebrated example being Newton's cradle) [17], [13], [6], [14]. Fibers assemblies [11], circuit breakers, systems with clearances, are also studied in the team.

4.6. Stability and Feedback Control

Lyapunov stability of nonsmooth, complementarity dynamical systems is challenging, because of possible state jumps, and varying system's dimension (the system may live on lower-dimensional subspaces), which may induce instability if not incorporated in the analysis [8], [9], [7]. On the other hand, the nonsmoothness (or the set-valuedness) may be introduced through the feedback control, like for instance the well-known sliding-mode controllers or state observers. The time-discretisation of set-valued controllers is in turn of big interest [3]. The techniques we study originate from numerical analysis in Contact Mechanics (the Moreau-Jean time-stepping algorithm) and are shown to be very efficient for chattering suppression and Lyapunov finite-time stability.

CAGIRE Project-Team

4. Application Domains

4.1. Aeroengines

Cagire is presently involved in studies mainly related with two subcomponents of the combustion chamber of aeroengines:

- The combustion chamber wall: the modelling, the simulation and the experimentation of the flow around a multiperforated plate representative of a real combustion chamber wall are the three axes we have been developing during the recent period. The continuous improvement of our in-house test facility Maveric is also an important ingredient to produce our own experimental validation data for isothermal flows. For non-isothermal flows, our participation in the EU funded program Soprano will be giving us access to non-isothermal data produced by Onera.

- The fuel nozzle: the impact of small geometrical variations related to the fabrication process on the pressure losses of the injector is being studied for a generic geometry provided by our industrial partner AD Industrie. RANS based simulations are intended to help designing some critical components of the fuel nozzle. A project is currently written to answer a call of the Nouvelle Aquitaine region, with the objective of extending this study to hybrid RANS/LES methods.

4.2. Power stations

R. Manceau has established a long term collaboration (4 CIFRE PhD theses in the past, 2 ongoing) with the R & D center of EDF of Chatou, for the development of refined turbulence models in the in-house CFD code of EDF, Code_Saturne :

- The prediction of heat transfer in fluid and solid components is of major importance in power stations, in particular, nuclear power plants. Either for the thermohydraulics of the plenum or in the study of accidental scenarii, among others, the accurate estimation of wall heat transfer, mean temperatures and temperature fluctuations are necessary for the evaluation of relevant thermal and mechanical design criteria. The PhD thesis (CIFRE EDF) of G. Mangeon is dedicated to the development of relevant RANS models for these industrial applications.

- Moreover, the prediction of unsteady hydrodynamic loadings is a key point for operating and for safety studies of PWR power plants. Currently, the static loading is correctly predicted by RANS computations but when the flow is transient (as, for instance, in Reactor Coolant Pumps, due to rotor/stator interactions, or during operating transients) or in the presence of large, energetic, coherent structures in the external flow region, the RANS approach is not sufficient, whereas LES is still too costly for a wide use in industry. This issue constitutes the starting point of the just-started PhD thesis (CIFRE EDF) of Vladimir Duffal.

4.3. Automotive propulsion

- The engine (underhood) compartment is a key component of vehicle design, in which the temperature is monitored to ensure the effectiveness and safety of the vehicle, and participates in 5 to 8% of the total drag and CO2 emissions. Dimensioning is an aerodynamic and aerothermal compromise, validated on a succession of road stages at constant speed and stopped phases (red lights, tolls, traffic jam). Although CFD is routinely used for forced convection, state-of-the-art turbulence models are not able to reproduce flows dominated by natural convection during stopped phases, with a Rayleigh number of the order of 10^{10}, such that the design still relies on costly, full-scale, wind tunnel experiments. This technical barrier must be lifted, since the ambition of the PSA group is to reach a *full digital design of their vehicles in the 2025 horizon*, i.e., to almost entirely rely on CFD. This issue is the focus of the ongoing PhD thesis (CIFRE PSA) of S. Jameel, supervised by R. Manceau, and also a part of the ANR project MONACO_2025 described in section 9.2.2 .

- The Power & Vehicles Division of IFPEN co-develops the CFD code CONVERGE to simulate the internal flow in a spark-ignition engine, in order to provide the automotive industry with tools to optimize the design of combustion engines. The RANS method, widely used in the industry, is not sufficiently reliable for quantitative predictions, and is only used as a tool to qualitatively compare different geometries. On the other hand, LES provides more detailed and accurate information, but at the price of a CPU cost unafordable for daily use in the industry. Therefore, IFPEN aims at developing the hybrid RANS/LES methodology, in order to combine the strengths of the two approaches. The just-started PhD thesis of Hassan Afailal, co-supervised by Rémi Manceau, is focused on this issue.

<p style="text-align: center; color: red; font-weight: bold;">CARDAMOM Project-Team</p>

4. Application Domains

4.1. De-anti icing systems

Impact of large ice debris on downstream aerodynamic surfaces and ingestion by aft mounted engines must be considered during the aircraft certification process. It is typically the result of ice accumulation on unprotected surfaces, ice accretions downstream of ice protected areas, or ice growth on surfaces due to delayed activation of ice protection systems (IPS) or IPS failure. This raises the need for accurate ice trajectory simulation tools to support pre-design, design and certification phases while improving cost efficiency. Present ice trajectory simulation tools have limited capabilities due to the lack of appropriate experimental aerodynamic force and moment data for ice fragments and the large number of variables that can affect the trajectories of ice particles in the aircraft flow field like the shape, size, mass, initial velocity, shedding location, etc... There are generally two types of model used to track shed ice pieces. The first type of model makes the assumption that ice pieces do not significantly affect the flow. The second type of model intends to take into account ice pieces interacting with the flow. We are concerned with the second type of models, involving fully coupled time-accurate aerodynamic and flight mechanics simulations, and thus requiring the use of high efficiency adaptive tools, and possibly tools allowing to easily track moving objects in the flow. We will in particular pursue and enhance our initial work based on adaptive immerse boundary capturing of moving ice debris, whose movements are computed using basic mechanical laws.

In [65] it has been proposed to model ice shedding trajectories by an innovative paradigm that is based on CArtesian grids, PEnalization and LEvel Sets (LESCAPE code). Our objective is to use the potential of high order unstructured mesh adaptation and immersed boundary techniques to provide a geometrically flexible extension of this idea. These activities will be linked to the development of efficient mesh adaptation and time stepping techniques for time dependent flows, and their coupling with the immersed boundary methods we started developing in the FP7 EU project STORM [55], [116]. In these methods we compensate for the error at solid walls introduced by the penalization by using anisotropic mesh adaptation [88], [106], [107]. From the numerical point of view one of the major challenges is to guarantee efficiency and accuracy of the time stepping in presence of highly stretched adaptive and moving meshes. Semi-implicit, locally implicit, multi-level, and split discretizations will be explored to this end.

Besides the numerical aspects, we will deal with modelling challenges. One source of complexity is the initial conditions which are essential to compute ice shedding trajectories. It is thus extremely important to understand the mechanisms of ice release. With the development of next generations of engines and aircraft, there is a crucial need to better assess and predict icing aspects early in design phases and identify breakthrough technologies for ice protection systems compatible with future architectures. When a thermal ice protection system is activated, it melts a part of the ice in contact with the surface, creating a liquid water film and therefore lowering ability of the ice block to adhere to the surface. The aerodynamic forces are then able to detach the ice block from the surface [67]. In order to assess the performance of such a system, it is essential to understand the mechanisms by which the aerodynamic forces manage to detach the ice. The current state of the art in icing codes is an empirical criterion. However such an empirical criterion is unsatisfactory. Following the early work of [71], [64] we will develop appropriate asymptotic PDE approximations allowing to describe the ice formation and detachment, trying to embed in this description elements from damage/fracture mechanics. These models will constitute closures for aerodynamics/RANS and URANS simulations in the form of PDE wall models, or modified boundary conditions.

In addition to this, several sources of uncertainties are associated to the ice geometry, size, orientation and the shedding location. In very few papers [120], some sensitivity analysis based on Monte Carlo method have been conducted to take into account the uncertainties of the initial conditions and the chaotic nature of the ice particle motion. We aim to propose some systematic approach to handle every source of uncertainty in an efficient way relying on some state-of-art techniques developed in the Team. In particular, we will perform an uncertainty propagation of some uncertainties on the initial conditions (position, orientation, velocity,...) through a low-fidelity model in order to get statistics of a multitude of particle tracks. This study will be done in collaboration with ETS (Ecole de Technologies Supérieure, Canada). The longterm objective is to produce footprint maps and to analyse the sensitivity of the models developed.

4.2. Space re-entry

As already mentioned, atmospheric re-entry involves multi-scale fluid flow physics including highly rarefied effects, aerothermochemistry, radiation. All this must be coupled to the response of thermal protection materials to extreme conditions. This response is most often the actual objective of the study, to allow the certification of Thermal Protection Systems (TPS).

One of the applications we will consider is the so-called post-flight analysis of a space mission. This involves reconstructing the history of the re-entry module (trajectory and flow) from data measured on the spacecraft by means of a Flush Air Data System (FADS), a set of sensors flush mounted in the thermal protection system to measure the static pressure (pressure taps) and heat flux (calorimeters). This study involves the accurate determination of the freestream conditions during the trajectory. In practice this means determining temperature, pressure, and Mach number in front of the bow shock forming during re-entry. As shown by zur Nieden and Olivier [145], state of the art techniques for freestream characterization rely on several approximations, such as e.g. using an equivalent calorically perfect gas formulas instead of taking into account the complex aero-thermo-chemical behaviour of the fluid. These techniques do not integrate measurement errors nor the heat flux contribution, for which a correct knowledge drives more complex models such as gas surface interaction. In this context, CFD supplied with UQ tools permits to take into account chemical effects and to include both measurement errors and epistemic uncertainties, e.g. those due to the fluid approximation, on the chemical model parameters in the bulk and at the wall (surface catalysis).

Rebuilding the freestream conditions from the stagnation point data therefore amounts to solving a stochastic inverse problem, as in robust optimization. Our objective is to build a robust and global framework for rebuilding freestream conditions from stagnation-point measurements for the trajectory of a re-entry vehicle. To achieve this goal, methods should be developed for

- an accurate simulation of the flow in all the regimes, from rarefied, to transitional, to continuous ;
- providing a complete analysis about the reliability and the prediction of the numerical simulation in hypersonic flows, determining the most important source of error in the simulation (PDE model, discretization, mesh, etc)
- reducing the overall computational cost of the analysis .

Our work on the improvement of the simulation capabilities for re-entry flows will focus both on the models and on the methods. We will in particular provide an approach to extend the use of standard CFD models in the transitional regime, with CPU gains of several orders of magnitude w.r.t. Boltzmann solvers. To do this we will use the results of a boundary layer analysis allowing to correct the Navier-Stokes equations. This theory gives modified (or extended) boundary conditions that are called "slip velocity" and "temperature jump" conditions. This theory seems to be completely ignored by the aerospace engineering community. Instead, people rather use a simpler theory due to Maxwell that also gives slip and jump boundary conditions: however, the coefficients given by this theory are not correct. This is why several teams have tried to modify these coefficients by some empirical methods, but it seems that this does not give any satisfactory boundary conditions.

Our project is twofold. First, we want to revisit the asymptotic theory, and to make it known in the aerospace community. Second, we want to make an intensive sensitivity analysis of the model to the various coefficients of the boundary conditions. Indeed, there are two kinds of coefficients in these boundary conditions. The first one is the accomodation coefficient: in the kinetic model, it gives the proportion of molecules that are specularly reflected, while the others are reflected according to a normal distribution (the so-called diffuse reflexion). This coefficient is a data of the kinetic model that can be measured by experiments: it depends on the material and the structure of the solid boundary, and of the gas. Its influence on the results of a Navier-Stokes simulation is certainly quite important. The other coefficients are those of the slip and jump boundary conditions: they are issued from the boundary layer analysis, and we have absolutely no idea of the order of magnitude of their influence on the results of a Navier-Stokes solution. In particular, it is not clear if these results are more sensitive to the accomodation coefficient or to these slip and jump coefficients.

In this project, we shall make use of the expertise of the team on uncertainty quantification to investigate the sensitivity of the Navier-Stokes model with slip and jump coefficients to these various coefficients. This would be rather new in the field of aerospace community. It could also have some impacts in other sciences in which slip and jump boundary conditions with incorrect coefficients are still used, like for instance in spray simulations: for very small particles immersed in a gas, the drag coefficient is modified to account for rarefied effects (when the radius of the particle is of the same order of magnitude as the mean free path in the gas), and slip and jump boundary conditions are used.

Another application which has very close similarities to the physics of de-anti icing systems is the modelling of the solid and liquid ablation of the thermal protective system of the aircraft. This involves the degradation and recession of the solid boundary of the protection layer due to the heating generated by the friction. As in the case of de-anti icing systems, the simulation of these phenomena need to take into account the heat conduction in the solid, its phase change, and the coupling between a weakly compressible and a compressible phase. Fluid/Solid coupling methods are generally based on a weak approach. Here we will both study, by theoretical and numerical techniques, a strong coupling method for the interaction between the fluid and the solid, and, as for de-anti icing systems, attempt at developing appropriate asymptotic models. These would constitute some sort of thin layer/wall models to couple to the external flow solver.

These modelling capabilities will be coupled to high order adaptive discretizations to provide high fidelity flow models. One of the most challenging problems is the minimization of the influence of mesh and scheme on the wall conditions on the re-entry module. To reduce this influence, we will investigate both high order adaptation across the bow shock, and possibly adaptation based on uncertainty quantification high order moments related to the heat flux estimation, or shock fitting techniques [68], [111]. These tools will be coupled to our robust inverse techniques. One of our objectives is to development of a low-cost strategy for improving the numerical prediction by taking into account experimental data. Some methods have been recently introduced [119] for providing an estimation of the numerical errors/uncertainties. We will use some metamodels for solving the inverse problem, by considering all sources of uncertainty, including those on physical models. We will validate the framework sing the experimental data available in strong collaboration with the von Karman Institute for Fluid dynamics (VKI). In particular, data coming from the VKI Longshot facility will be used. We will show application of the developed numerical tool for the prediction in flight conditions.

These activities will benefit from our strong collaborations with the CEA and with the von Karman Institute for Fluid Dynamics and ESA.

4.3. Energy

We will develop modelling and design tools, as well as dedicated platforms, for Rankine cycles using complex fluids (organic compounds), and for wave energy extraction systems.

Organic Rankine Cycles (ORCs) use heavy organic compounds as working fluids. This results in superior efficiency over steam Rankine cycles for source temperatures below 900 K. ORCs typically require only a single-stage rotating component making them much simpler than typical multi-stage steam turbines. The strong pressure reduction in the turbine may lead to supersonic flows in the rotor, and thus to the appearance of shocks, which reduces the efficiency due to the associated losses. To avoid this, either a larger multi stage installation is used, in which smaller pressure drops are obtained in each stage, or centripetal turbines are used, at very high rotation speeds (of the order of 25,000 rpm). The second solution allows to keep the simplicity of the expander, but leads to poor turbine efficiencies (60-80%) - w.r.t. modern, highly optimized, steam and gas turbines - and to higher mechanical constraints. The use of *dense-gas working fluids*, *i.e.* operating close to the saturation curve, in properly chosen conditions could increase the turbine critical Mach number avoiding the formation of shocks, and increasing the efficiency. Specific shape optimization may enhance these effects, possibly allowing the reduction of rotation speeds. However, dense gases may have significantly different properties with respect to dilute ones. Their dynamics is governed by a thermodynamic parameter known as the fundamental derivative of gas dynamics

$$\Gamma = 1 + \frac{\rho}{c} \left(\frac{\partial c}{\partial \rho} \right)_s , \tag{2}$$

where ρ is the density, c is the speed of sound and s is the entropy. For ideal gas $\Gamma = (\gamma + 1)/2 > 1$. For some complex fluids and some particular conditions of pressure and temperature, Γ may be lower that one, implying that $(\partial c/\partial \rho)_s < 0$. This means that the acceleration of pressure perturbations through a variable density fluids may be reversed and become a deceleration. It has been shown that, for $\Gamma << 1$, compression shocks are strongly reduced, thus alleviating the shock intensity. This has great potential in increasing the efficiency. This is why so much interest is put on dense gas ORCs.

The simulation of these gases requires accurate thermodynamic models, such as Span-Wagner or Peng-Robinson (see [79]). The data to build these models is scarce due to the difficulty of performing reliable experiments. The related uncertainty is thus very high. Our work will go in the following directions:

1. develop deterministic models for the turbine and the other elements of the cycle. These will involve multi-dimensional high fidelity, as well as intermediate and low fidelity (one- and zero-dimensional), models for the turbine, and some 0D/1D models for other element of the cycle (pump, condenser, etc) ;

2. validation of the coupling between the various elements. The following aspects will be considered: characterization of the uncertainties on the cycle components (e.g. empirical coefficients modelling the pump or the condenser), calibration of the thermodynamic parameters, model the uncertainty of each element, and the influence of the unsteady experimental data ;

3. demonstrate the interest of a specific optimization of geometry, operating conditions, and the choice of the fluid, according to the geographical location by including local solar radiation data. Multi-objective optimization will be considered to maximize performance indexes (e.g. Carnot efficiency, mechanical work and energy production), and to reduce the variability of the output.

This work will provide modern tools for the robust design of ORCs systems. It benefits from the direct collaboration with the SME EXOES (ANR LAbCom VIPER), and from a collaboration with LEMMA.
Wave energy conversion is an emerging sector in energy engineering. The design of new and efficient Wave Energy Converters (WECs) is thus a crucial activity. As pointed out by Weber [143], it is more economical to raise the technology performance level (TPL) of a wave energy converter concept at low technology readiness level (TRL). Such a development path puts a greater demand on the numerical methods used. The findings of Weber also tell us that important design decisions as well as optimization should be performed as early in the development process as possible. However, as already mentioned, today the wave energy sector relies heavily on the use of tools based on simplified linear hydrodynamic models for the prediction of motions, loads, and power production. Our objective is to provide this sector, and especially SMEs, with robust design tools to minimize the uncertainties in predicted power production, loads, and costs of wave energy.

Following our initial work [91], we will develop, analyse, compare, and use for multi-fidelity optimization,non-linear models of different scales (fidelity) ranging from simple linear hydrodynamics over asymptotic discrete nonlinear wave models, to non-hydrostatic anisoptropic Euler free surface solvers. We will not work on the development of small scale models (VOF-RANS or LES) but may use such models, developed by our collaborators, for validation purposes. These developments will benefit from all our methodological work on asymptotic modelling and high order discretizations. As shown in [91], asymptotic models foe WECs involve an equation for the pressure on the body inducing a PDE structure similar to that of incompressible flow equations. The study of appropriate stable and efficient high order approximations (coupling velocity-pressure, efficient time stepping) will be an important part of this activity. Moreover, the flow-floating body interaction formulation introduces time stepping issues similar to those encountered in fluid structure interaction problems, and require a clever handling of complex floater geometries based on adaptive and ALE techniques. For this application, the derivation of fully discrete asymptotics may actually simplify our task.

Once available, we will use this hierarchy of models to investigate and identify the modelling errors, and provide a more certain estimate of the cost of wave energy. Subsequently we will look into optimization cycles by comparing time-to-decision in a multi-fidelity optimization context. In particular, this task will include the development and implementation of appropriate surrogate models to reduce the computational cost of expensive high fidelity models. Here especially artificial neural networks (ANN) and Kriging response surfaces (KRS) will be investigated. This activity on asymptotic non-linear modelling for WECs, which has had very little attention in the past, will provide entirely new tools for this application. Multi-fidelity robust optimization is also an approach which has never been applied to WECs.

This work is the core of the EU OCEANEranet MIDWEST project, which we coordinate. It will be performed in collaboration with our European partners, and with a close supervision of European SMEs in the sector, which are part of the steering board of MIDWEST (WaveDragon, Waves4Power, Tecnalia).

4.4. Materials engineering

Because of their high strength and low weight, ceramic-matrix composite materials (CMCs) are the focus of active research for aerospace and energy applications involving high temperatures, either military or civil. Though based on brittle ceramic components, these composites are not brittle due to the use of a fibre/matrix interphase that preserves the fibres from cracks appearing in the matrix. Recent developments aim at implementing also in civil aero engines a specific class of Ceramic Matrix Composite materials (CMCs) that show a self-healing behaviour. Self-healing consists in filling cracks appearing in the material with a dense fluid formed in-situ by oxidation of part of the matrix components. Self-healing (SH) CMCs are composed of a complex three-dimensional topology of woven fabrics containing fibre bundles immersed in a matrix coating of different phases. The oxide seal protects the fibres which are sensitive to oxidation, thus delaying failure. The obtained lifetimes reach hundreds of thousands of hours [123].

The behaviour of a fibre bundle is actually extremely variable, as the oxidation reactions generating the self-healing mechanism have kinetics strongly dependent on temperature and composition. In particular, the lifetime of SH-CMCs depends on: (i) temperature and composition of the surrounding atmosphere; (ii) composition and topology of the matrix layers; (iii) the competition of the multidimensional diffusion/oxidation/volatilization processes; (iv) the multidimensional flow of the oxide in the crack; (v) the inner topology of fibre bundles; (vi) the distribution of critical defects in the fibres. Unfortunately, experimental investigations on the full materials are too long (they can last years) and their output too qualitative (the coupled effects can only be observed a-posteriori on a broken sample). Modelling is thus essential to study and to design SH-CMCs.

In collaboration wit the LCTS laboratory (a joint CNRS-CEA-SAFRAN-Bordeaux University lab devoted to the study of thermo-structural materials in Bordeaux), we are developing a multi-scale model in which a structural mechanics solver is coupled with a closure model for the crack physico chemistry. This model is obtained as a multi-dimensional asymptotic crack averaged approximation fo the transport equations (Fick's laws) with chemical reactions sources, plus a potential model for the flow of oxide [82], [89], [121]. We

have demonstrated the potential of this model in showing the importance of taking into account the multi-dimensional topology of a fibre bundle (distribution of fibres) in the rupture mechanism. This means that the 0-dimensional model used in most of the studies (se e.g. [78]) will underestimate appreciably the lifetime of the material. Based on these recent advances, we will further pursue the development of multi-scale multi-dimensional asymptotic closure models for the parametric design of self healing CMCs. Our objectives are to provide: (i) new, non-linear multi-dimensional mathematical model of CMCs, in which the physico-chemistry of the self-healing process is more strongly coupled to the two-phase (liquid gas) hydro-dynamics of the healing oxide ; (ii) a model to represent and couple crack networks ; (iii) a robust and efficient coupling with the structural mechanics code ; (iv) validate this platform with experimental data obtained at the LCTS laboratory. The final objective is to set up a multi-scale platform for the robust prediction of lifetime of SH-CMCs, which will be a helpful tool for the tailoring of the next generation of these materials.

4.5. Coastal and civil engineering

Our objective is to bridge the gap between the development of high order adaptive methods, which has mainly been performed in the industrial context and environmental applications, with particular attention to coastal and hydraulic engineering. We want to provide tools for adaptive non-linear modelling at large and intermediate scales (near shore, estuarine and river hydrodynamics). We will develop multi-scale adaptive models for free surface hydrodynamics. Beside the models and codes themselves, based on the most advanced numerics we will develop during this project, we want to provide sufficient know how to control, adapt and optimize these tools.

We will focus our effort in the understanding of the interactions between asymptotic approximations and numerical approximations. This is extremely important in at least two aspects. The first is the capability of a numerical model to handle highly dispersive wave propagation. This is usually done by high accuracy asymptotic PDE expansions. Here we plan to make heavily use of our results concerning the relations between vertical asymptotic expansions and standard finite element approximations. In particular, we will invest some effort in the development of $xy+z$ adaptive finite element approximations of the incompressible Euler equations. Local $p-$adaptation of the vertical approximation may provide a "variable depth" approximation exploiting numerics instead of analytical asymptotics to control the physical behaviour of the model.

Another important aspect which is not understood well enough at the moment is the role of dissipation in wave breaking regions. There are several examples of breaking closure, going from algebraic and PDE-based eddy viscosity methods [103], [127], [118], [86], to hybrid methods coupling dispersive PDEs with hyperbolic ones, and trying to mimic wave breaking with travelling bores [136], [137], [135], [101], [93]. In both cases, numerical dissipation plays an important role and the activation or not of the breaking closure, as the quantitative contribution of numerical dissipation to the flow has not been properly investigated. These elements must be clarified to allow full control of adaptive techniques for the models used in this type of applications.

Another point we want to clarify is how to optimize the discretization of asymptotic PDE models. In particular, when adding mesh size(s) and time step, we are in presence of at least 3 (or even more) small parameters. The relations between physical ones have been more or less investigates, as have been the ones between purely numerical ones. We plan to study the impact of numerics on asymptotic PDE modelling by reverting the usual process and studying asymptotic limits of finite element discretizations of the Euler equations. Preliminary results show that this does allow to provide some understanding of this interaction and to possibly propose considerably improved numerical methods [66].

COMMANDS Project-Team

4. Application Domains

4.1. Fuel saving by optimizing airplanes trajectories

We have a collaboration with the startup Safety Line on the optimization of trajectories for civil aircrafts. Key points include the reliable identification of the plane parameters (aerodynamic and thrust models) using data from the flight recorders, and the robust trajectory optimization of the climbing and cruise phases. We use both local (quasi-Newton interior-point algorithms) and global optimization tools (dynamic programming). The local method for the climb phase is in production and has been used for several hundreds of actual plane flights.

4.2. Hybrid vehicles

We have a collaboration with IFPEN on the energy management for hybrid vehicles. A significant direction is the analysis and classification of traffic data. More specifically, we focus on the traffic probability distribution in the (speed,torque) plane, with a time / space subdivision (road segments and timeframes).

4.3. Biological systems

We renewed in 2017 our interest in (micro)biological systems, joining projects Cosy and Algae in silico on the topic of the optimization of micro-organisms populations.

CQFD Project-Team

4. Application Domains

4.1. Dependability and safety

Our abilities in probability and statistics apply naturally to industry, in particular in studies of dependability and safety. An illustrative example is the collaboration that started in September 2014 with with THALES Optronique. The goal of this project is the optimization of the maintenance of an onboard system equipped with a HUMS (Health Unit Monitoring Systems). The physical system under consideration is modeled by a piecewise deterministic Markov process. In the context of impulse control, we propose a dynamic maintenance policy, adapted to the state of the system and taking into account both random failures and those related to the degradation phenomenon.

The spectrum of applications of the topics that the team can address is large and can concern many other fields. Indeed non parametric and semi-parametric regression methods can be used in biometry, econometrics or engineering for instance. Gene selection from microarray data and text categorization are two typical application domains of dimension reduction among others. We had for instance the opportunity via the scientific program PRIMEQUAL to work on air quality data and to use dimension reduction techniques as principal component analysis (PCA) or positive matrix factorization (PMF) for pollution sources identification and quantization.

<p style="color:red; text-align:center;">**DEFI Project-Team**</p>

4. Application Domains

4.1. Radar and GPR applications

Conventional radar imaging techniques (ISAR, GPR, etc.) use backscattering data to image targets. The commonly used inversion algorithms are mainly based on the use of weak scattering approximations such as the Born or Kirchhoff approximation leading to very simple linear models, but at the expense of ignoring multiple scattering and polarization effects. The success of such an approach is evident in the wide use of synthetic aperture radar techniques.

However, the use of backscattering data makes 3-D imaging a very challenging problem (it is not even well understood theoretically) and as pointed out by Brett Borden in the context of airborne radar: "In recent years it has become quite apparent that the problems associated with radar target identification efforts will not vanish with the development of more sensitive radar receivers or increased signal-to-noise levels. In addition it has (slowly) been realized that greater amounts of data - or even additional "kinds" of radar data, such as added polarization or greatly extended bandwidth - will all suffer from the same basic limitations affiliated with incorrect model assumptions. Moreover, in the face of these problems it is important to ask how (and if) the complications associated with radar based automatic target recognition can be surmounted." This comment also applies to the more complex GPR problem.

Our research themes will incorporate the development, analysis and testing of several novel methods, such as sampling methods, level set methods or topological gradient methods, for ground penetrating radar application (imaging of urban infrastructures, landmines detection, underground waste deposits monitoring,) using multistatic data.

4.2. Biomedical imaging

Among emerging medical imaging techniques we are particularly interested in those using low to moderate frequency regimes. These include Microwave Tomography, Electrical Impedance Tomography and also the closely related Optical Tomography technique. They all have the advantage of being potentially safe and relatively cheap modalities and can also be used in complementarity with well established techniques such as X-ray computed tomography or Magnetic Resonance Imaging.

With these modalities tissues are differentiated and, consequentially can be imaged, based on differences in dielectric properties (some recent studies have proved that dielectric properties of biological tissues can be a strong indicator of the tissues functional and pathological conditions, for instance, tissue blood content, ischemia, infarction, hypoxia, malignancies, edema and others). The main challenge for these functionalities is to built a 3-D imaging algorithm capable of treating multi-static measurements to provide real-time images with highest (reasonably) expected resolutions and in a sufficiently robust way.

Another important biomedical application is brain imaging. We are for instance interested in the use of EEG and MEG techniques as complementary tools to MRI. They are applied for instance to localize epileptic centers or active zones (functional imaging). Here the problem is different and consists into performing passive imaging: the epileptic centers act as electrical sources and imaging is performed from measurements of induced currents. Incorporating the structure of the skull is primordial in improving the resolution of the imaging procedure. Doing this in a reasonably quick manner is still an active research area, and the use of asymptotic models would offer a promising solution to fix this issue.

4.3. Non destructive testing and parameter identification

One challenging problem in this vast area is the identification and imaging of defaults in anisotropic media. For instance this problem is of great importance in aeronautic constructions due to the growing use of composite materials. It also arises in applications linked with the evaluation of wood quality, like locating knots in timber in order to optimize timber-cutting in sawmills, or evaluating wood integrity before cutting trees. The anisotropy of the propagative media renders the analysis of diffracted waves more complex since one cannot only relies on the use of backscattered waves. Another difficulty comes from the fact that the micro-structure of the media is generally not well known a priori.

Our concern will be focused on the determination of qualitative information on the size of defaults and their physical properties rather than a complete imaging which for anisotropic media is in general impossible. For instance, in the case of homogeneous background, one can link the size of the inclusion and the index of refraction to the first eigenvalue of so-called interior transmission problem. These eigenvalues can be determined form the measured data and a rough localization of the default. Our goal is to extend this kind of idea to the cases where both the propagative media and the inclusion are anisotropic. The generalization to the case of cracks or screens has also to be investigated.

In the context of nuclear waste management many studies are conducted on the possibility of storing waste in a deep geological clay layer. To assess the reliability of such a storage without leakage it is necessary to have a precise knowledge of the porous media parameters (porosity, tortuosity, permeability, etc.). The large range of space and time scales involved in this process requires a high degree of precision as well as tight bounds on the uncertainties. Many physical experiments are conducted in situ which are designed for providing data for parameters identification. For example, the determination of the damaged zone (caused by excavation) around the repository area is of paramount importance since microcracks yield drastic changes in the permeability. Level set methods are a tool of choice for characterizing this damaged zone.

4.4. Diffusion MRI

In biological tissues, water is abundant and magnetic resonance imaging (MRI) exploits the magnetic property of the nucleus of the water proton. The imaging contrast (the variations in the grayscale in an image) in standard MRI can be from either proton density, T1 (spin-lattice) relaxation, or T2 (spin-spin) relaxation and the contrast in the image gives some information on the physiological properties of the biological tissue at different physical locations of the sample. The resolution of MRI is on the order of millimeters: the greyscale value shown in the imaging pixel represents the volume-averaged value taken over all the physical locations contained that pixel.

In diffusion MRI, the image contrast comes from a measure of the average distance the water molecules have moved (diffused) during a certain amount of time. The Pulsed Gradient Spin Echo (PGSE) sequence is a commonly used sequence of applied magnetic fields to encode the diffusion of water protons. The term 'pulsed' means that the magnetic fields are short in duration, an the term gradient means that the magnetic fields vary linearly in space along a particular direction. First, the water protons in tissue are labelled with nuclear spin at a precession frequency that varies as a function of the physical positions of the water molecules via the application of a pulsed (short in duration, lasting on the order of ten milliseconds) magnetic field. Because the precessing frequencies of the water molecules vary, the signal, which measures the aggregate phase of the water molecules, will be reduced due to phase cancellations. Some time (usually tens of milliseconds) after the first pulsed magnetic field, another pulsed magnetic field is applied to reverse the spins of the water molecules. The time between the applications of two pulsed magnetic fields is called the 'diffusion time'. If the water molecules have not moved during the diffusion time, the phase dispersion will be reversed, hence the signal loss will also be reversed, the signal is called refocused. However, if the molecules have moved during the diffusion time, the refocusing will be incomplete and the signal detected by the MRI scanner if weaker than if the water molecules have not moved. This lack of complete refocusing is called the signal attenuation and is the basis of the image contrast in DMRI. the pixels showning more signal attenuation is associated with further water displacement during the diffusion time, which may be linked to physiological factors, such as higher cell membrane permeability, larger cell sizes, higher extra-cellular volume fraction.

We model the nuclear magnetization of water protons in a sample due to diffusion-encoding magnetic fields by a multiple compartment Bloch-Torrey partial differential equation, which is a diffusive-type time-dependent PDE. The DMRI signal is the integral of the solution of the Bloch-Torrey PDE. In a homogeneous medium, the intrinsic diffusion coeffcient D will appear as the slope of the semi-log plot of the signal (in appropriate units). However, because during typical scanning times, 50-100ms, water molecules have had time to travel a diffusion distance which is long compared to the average size of the cells, the slope of the semi-log plot of the signal is in fact a measure of an 'effective' diffusion coefficient. In DMRI applications, this measured quantity is called the 'apparent diffusion coefficient' (ADC) and provides the most commonly used form the image contrast for DMRI. This ADC is closely related to the effective diffusion coefficient obtainable from mathematical homogenization theory.

DISCO Project-Team

4. Application Domains

4.1. Analysis and Control of life sciences systems

The team is involved in life sciences applications. The two main lines are the analysis of bioreactors models and the modeling of cell dynamics in Acute Myeloblastic Leukemias (AML) in collaboration with St Antoine Hospital in Paris. A recent new subject is the modelling of Dengue epidemics.

4.2. Energy Management

The team is interested in Energy management and considers optimization and control problems in energy networks.

DOLPHIN Team

4. Application Domains

4.1. Smart grids

With the smart grid revolution, house energy consumption will play a significant role in the energy system. Home users are indeed responsible for a significant portion of the world's energy needs portion, but are totally inelastic with respect to the market (i.e. the energy demand does not follow the price of the energy itself). Thus, the whole energy generation and distribution system performance can be improved by optimizing the house energy management. Those problems are concerned by multiple objectives such as cost and users' comfort, and multiple decision makers such as end-users and energy operators. We propose a home automation system that can monitor appliance scheduling in order to simultaneously optimize the total energy cost and the customer satisfaction [17].

The key challenge is to propose new optimization models and new hybrid optimization algorithms to the demand side management of smart grids in a context of uncertainty and in the presence of several conflicting objectives. Those complex optimization problems are also characterized by the presence of both continuous and discrete variables [18].

4.2. Transportation and logistics

- **Scheduling problems under uncertainty:** The flow-shop scheduling problem is one of the most well-known problems from scheduling. However, most of the works in the literature use a deterministic single-objective formulation. In general, the minimized objective is the total completion time (makespan). Many other criteria may be used to schedule tasks on different machines: maximum tardiness, total tardiness, mean job flowtime, number of delayed jobs, maximum job flowtime, etc. In the DOLPHIN project, a bi-criteria model, which consists in minimizing the makespan and the total tardiness, is studied. A bi-objective flow-shop problem with uncertainty on the duration, minimizing in addition the maximum tardiness, is also studied. It allows us to develop and test multi-objective (and not only bi-objective) optimization methods under uncertainty.

- **Routing problems under uncertainty:** The vehicle routing problem (VRP) is a well-known problem and it has been studied since the end of the fifties. It has a lot of practical applications in many industrial areas (ex. transportation, logistics, etc). Existing studies of the VRP are almost all concerned with the minimization of the total distance only. The model studied in the DOLPHIN project introduces a second objective, whose purpose is to balance the length of the tours. This new criterion is expressed as the minimization of the difference between the length of the longest tour and the length of the shortest tour. Uncertainty on the demands has also been introduced in the model.

4.3. Bioinformatics and Health care

Bioinformatic research is a great challenge for our society and numerous research entities of different specialities (biology, medical or information technology) are collaborating on specific themes.

4.3.1. *Genomic and post-genomic studies*

Previous studies of the DOLPHIN project mainly deal with genomic and postgenomic applications. These have been realized in collaboration with academic and industrial partners (IBL: Biology Institute of Lille; IPL: Pasteur Institute of Lille; IT-Omics firm).

First, genomic studies aim at analyzing genetic factors which may explain multi-factorial diseases such as diabetes, obesity or cardiovascular diseases. The scientific goal was to formulate hypotheses describing associations that may have any influence on diseases under study.

Secondly, in the context of post-genomic, a very large amount of data are obtained thanks to advanced technologies and have to be analyzed. Hence, one of the goals of the project was to develop analysis methods in order to discover knowledge in data coming from biological experiments.

These problems can be modeled as classical data mining tasks (Association rules, feature selection). As the combinatoric of such problems is very high and the quality criteria not unique, we proposed to model these problems as multi-objective combinatorial optimization problems. Evolutionary approaches have been adopted in order to cope with large scale problems.

Nowadays the technology is still going fast and the amount of data increases rapidly. Within the collaboration with Genes Diffusion, specialized in genetics and animal reproduction for bovine, swine, equine and rabbit species, we study combinations of Single Nucleotide Polymorphisms (SNP) that can explain some phenotypic characteristics. Therefore feature selection for regression is addressed using metaheuristics.

4.3.2. *Optimization for health care*

The collaboration with the Alicante company, a major actor in the hospital decision making, deals with knowledge extraction by optimization methods for improving the process of inclusion in clinical trials. Indeed, conducting a clinical trial, allowing for example to measure the effectiveness of a treatment, involves selecting a set of patients likely to participate to this test. Currently existing selection processes are far from optimal, and many potential patients are not considered. The objective of this collaboration consists in helping the practitioner to quickly determine if a patient is interesting for a clinical trial or not. Exploring different data sources (from a hospital information system, patient data...), a set of decision rules have to be generated. For this, approaches from multi-objective combinatorial optimization are implemented, requiring extensive work to model the problem, to define criteria optimization and to design specific optimization methods.

In another project we address the problems, proper models and suitable solving approaches to optimize the design and planning of a medical laboratory. It comes from the observation that human biology laboratories tend to get bigger and more complicated. The goal of this project is to design and apply solutions to optimize human biology laboratory decisions and operations. As a result of mutualisation phenomenon, huge medical laboratories are emerging all over the world. Medical laboratories are responsible for analyzing medical tests ordered by physicians on patient's samples. Large number of prescriptions and tubes are received by a laboratory make a complicated workflow into the system. The aim of this thesis is to design and plan a medical laboratory to minimize the costs and time required to perform the tests. Medical laboratory optimization is an immense problem which encompasses many subproblems. In the literature, there are only few studies that refer directly to medical laboratory planning and optimization from different points of view. In this thesis, the problem of designing and planning of medical laboratory is studied in a comprehensive form for both existing and new labs to provide practical models and solutions that can be implemented to real cases. The problem is studied from two different aspects: strategic and operational problems. In strategic level, machine selection and analyzer configuration problem and facilities layout problem are modeled and solved to achieve an optimal medical laboratory design. Mathematical programming is used for this phase. For operational level, the problems of assignment and scheduling are emerged. Mathematical programming, heuristic and meta-heuristic algorithms are proposed to deal with such problems. Simulation as a powerful tool is applied to verify and validate the solutions proposed from the previous analytical steps and also to aid decision making in some problems0. FlexSim is a simulation software which is used in this study and provide us the ability to model and analyze complex systems. It is worth mentioning that this project is raised by Normand-Info, a Beckman Coulter subsidiary which is an international company mainly producing human biology instruments. Thus, it is a collaborative research work with industry.

<p style="text-align:center">**ECUADOR Project-Team**</p>

4. Application Domains

4.1. Algorithmic Differentiation

Algorithmic Differentiation of programs gives sensitivities or gradients, useful for instance for :

- optimum shape design under constraints, multidisciplinary optimization, and more generally any algorithm based on local linearization,
- inverse problems, such as parameter estimation and in particular 4Dvar data assimilation in climate sciences (meteorology, oceanography),
- first-order linearization of complex systems, or higher-order simulations, yielding reduced models for simulation of complex systems around a given state,
- mesh adaptation and mesh optimization with gradients or adjoints,
- equation solving with the Newton method,
- sensitivity analysis, propagation of truncation errors.

4.2. Multidisciplinary optimization

A CFD program computes the flow around a shape, starting from a number of inputs that define the shape and other parameters. On this flow one can define optimization criteria e.g. the lift of an aircraft. To optimize a criterion by a gradient descent, one needs the gradient of the criterion with respect to all inputs, and possibly additional gradients when there are constraints. Adjoint AD is the most efficient way to compute these gradients.

4.3. Inverse problems and Data Assimilation

Inverse problems aim at estimating the value of hidden parameters from other measurable values, that depend on the hidden parameters through a system of equations. For example, the hidden parameter might be the shape of the ocean floor, and the measurable values of the altitude and velocities of the surface. Figure 1 shows an example of an inverse problem using the glaciology code ALIF (a pure C version of ISSM [29]) and its AD-adjoint produced by Tapenade.

One particular case of inverse problems is *data assimilation* [30] in weather forecasting or in oceanography. The quality of the initial state of the simulation conditions the quality of the prediction. But this initial state is not well known. Only some measurements at arbitrary places and times are available. A good initial state is found by solving a least squares problem between the measurements and a guessed initial state which itself must verify the equations of meteorology. This boils down to solving an adjoint problem, which can be done though AD [32]. The special case of *4Dvar* data assimilation is particularly challenging. The 4th dimension in "4D" is time, as available measurements are distributed over a given assimilation period. Therefore the least squares mechanism must be applied to a simulation over time that follows the time evolution model. This process gives a much better estimation of the initial state, because both position and time of measurements are taken into account. On the other hand, the adjoint problem involved is more complex, because it must run (backwards) over many time steps. This demanding application of AD justifies our efforts in reducing the runtime and memory costs of AD adjoint codes.
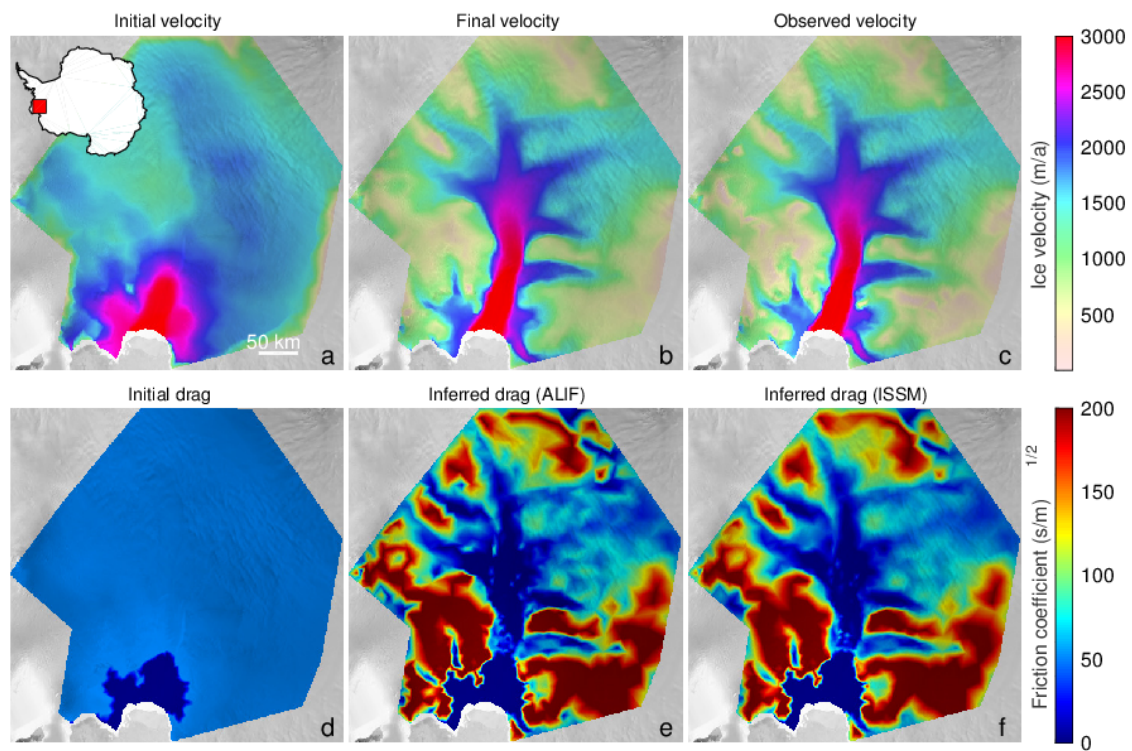
Figure 1. Assimilation of the basal friction under Pine Island glacier, West Antarctica. The final simulated surface velocity (b) is made to match the observed surface velocity (c), by estimation of the basal friction (e). A reference basal friction (f) is obtained by another data assimilation using the hand=written adjoint of ISSM

4.4. Linearization

Simulating a complex system often requires solving a system of Partial Differential Equations. This can be too expensive, in particular for real-time simulations. When one wants to simulate the reaction of this complex system to small perturbations around a fixed set of parameters, there is an efficient approximation: just suppose that the system is linear in a small neighborhood of the current set of parameters. The reaction of the system is thus approximated by a simple product of the variation of the parameters with the Jacobian matrix of the system. This Jacobian matrix can be obtained by AD. This is especially cheap when the Jacobian matrix is sparse. The simulation can be improved further by introducing higher-order derivatives, such as Taylor expansions, which can also be computed through AD. The result is often called a *reduced model*.

4.5. Mesh adaptation

Some approximation errors can be expressed by an adjoint state. Mesh adaptation can benefit from this. The classical optimization step can give an optimization direction not only for the control parameters, but also for the approximation parameters, and in particular the mesh geometry. The ultimate goal is to obtain optimal control parameters up to a precision prescribed in advance.

GAMMA3 Project-Team (section vide)

GECO Project-Team

4. Application Domains

4.1. Quantum control

The issue of designing efficient transfers between different atomic or molecular levels is crucial in atomic and molecular physics, in particular because of its importance in those fields such as photochemistry (control by laser pulses of chemical reactions), nuclear magnetic resonance (NMR, control by a magnetic field of spin dynamics) and, on a more distant time horizon, the strategic domain of quantum computing. This last application explicitly relies on the design of quantum gates, each of them being, in essence, an open loop control law devoted to a prescribed simultaneous control action. NMR is one of the most promising techniques for the implementation of a quantum computer.

Physically, the control action is realized by exciting the quantum system by means of one or several external fields, being them magnetic or electric fields. The resulting control problem has attracted increasing attention, especially among quantum physicists and chemists (see, for instance, [68], [73]). The rapid evolution of the domain is driven by a multitude of experiments getting more and more precise and complex (see the recent review [29]). Control strategies have been proposed and implemented, both on numerical simulations and on physical systems, but there is still a large gap to fill before getting a complete picture of the control properties of quantum systems. Control techniques should necessarily be innovative, in order to take into account the physical peculiarities of the model and the specific experimental constraints.

The area where the picture got clearer is given by finite dimensional linear closed models.

- **Finite dimensional** refers to the dimension of the space of wave functions, and, accordingly, to the finite number of energy levels.
- **Linear** means that the evolution of the system for a fixed (constant in time) value of the control is determined by a linear vector field.
- **Closed** refers to the fact that the systems are assumed to be totally disconnected from the environment, resulting in the conservation of the norm of the wave function.

The resulting model is well suited for describing spin systems and also arises naturally when infinite dimensional quantum systems of the type discussed below are replaced by their finite dimensional Galerkin approximations. Without seeking exhaustiveness, let us mention some of the issues that have been tackled for finite dimensional linear closed quantum systems:

- controllability [11],
- bounds on the controllability time [7],
- STIRAP processes [78],
- simultaneous control [51],
- optimal control ([47], [20], [31]),
- numerical simulations [57].

Several of these results use suitable transformations or approximations (for instance the so-called rotating wave) to reformulate the finite-dimensional Schrödinger equation as a sub-Riemannian system. Open systems have also been the object of an intensive research activity (see, for instance, [12], [48], [69], [26]).

In the case where the state space is infinite dimensional, some optimal control results are known (see, for instance, [16], [27], [44], [17]). The controllability issue is less understood than in the finite dimensional setting, but several advances should be mentioned. First of all, it is known that one cannot expect exact controllability on the whole Hilbert sphere [77]. Moreover, it has been shown that a relevant model, the quantum oscillator, is not even approximately controllable [70], [60]. These negative results have been more recently completed by positive ones. In [18], [19] Beauchard and Coron obtained the first positive controllability result for a quantum particle in a 1D potential well. The result is highly nontrivial and is based on Coron's return method (see [33]). Exact controllability is proven to hold among regular enough wave functions. In particular, exact controllability among eigenfunctions of the uncontrolled Schrödinger operator can be achieved. Other important approximate controllability results have then been proved using Lyapunov methods [59], [64], [45]. While [59] studies a controlled Schrödinger equation in $\mathbb{R}$ for which the uncontrolled Schrödinger operator has mixed spectrum, [64], [45] deal mainly with general discrete-spectrum Schrödinger operators.

In all the positive results recalled in the previous paragraph, the quantum system is steered by a single external field. Different techniques can be applied in the case of two or more external fields, leading to additional controllability results [36], [23].

The picture is even less clear for nonlinear models, such as Gross–Pitaevski and Hartree–Fock equations. The obstructions to exact controllability, similar to the ones mentioned in the linear case, have been discussed in [42]. Optimal control approaches have also been considered [15], [28]. A comprehensive controllability analysis of such models is probably a long way away.

4.2. Neurophysiology

At the interface between neurosciences, mathematics, automatics and humanoid robotics, an entire new approach to neurophysiology is emerging. It arouses a strong interest in the four communities and its development requires a joint effort and the sharing of complementary tools.

A family of extremely interesting problems concerns the understanding of the mechanisms supervising some sensorial reactions or biomechanics actions such as image reconstruction by the primary visual cortex, eyes movement and body motion.

In order to study these phenomena, a promising approach consists in identifying the motion planning problems undertaken by the brain, through the analysis of the strategies that it applies when challenged by external inputs. The role of control is that of a language allowing to read and model neurological phenomena. The control algorithms would shed new light on the brain's geometric perception (the so-called neurogeometry [66]) and on the functional organization of the motor pathways.

- A challenging problem is that of the understanding of the mechanisms which are responsible for the process of image reconstruction in the primary visual cortex V1.

  The visual cortex areas composing V1 are notable for their complex spatial organization and their functional diversity. Understanding and describing their architecture requires sophisticated modeling tools. At the same time, the structure of the natural and artificial images used in visual psychophysics can be fully disclosed only using rather deep geometric concepts. The word "geometry" refers here to the internal geometry of the functional architecture of visual cortex areas (not to the geometry of the Euclidean external space). Differential geometry and analysis both play a fundamental role in the description of the structural characteristics of visual perception.

  A model of human perception based on a simplified description of the visual cortex V1, involving geometric objects typical of control theory and sub-Riemannian geometry, has been first proposed by Petitot ([67]) and then modified by Citti and Sarti ([32]). The model is based on experimental observations, and in particular on the fundamental work by Hubel and Wiesel [41] who received the Nobel prize in 1981.

In this model, neurons of V1 are grouped into orientation columns, each of them being sensitive to visual stimuli arriving at a given point of the retina and oriented along a given direction. The retina is modeled by the real plane, while the directions at a given point are modeled by the projective line. The fiber bundle having as base the real plane and as fiber the projective line is called the *bundle of directions of the plane*.

From the neurological point of view, orientation columns are in turn grouped into hypercolumns, each of them sensitive to stimuli arriving at a given point, oriented along any direction. In the same hypercolumn, relative to a point of the plane, we also find neurons that are sensitive to other stimuli properties, such as colors. Therefore, in this model the visual cortex treats an image not as a planar object, but as a set of points in the bundle of directions of the plane. The reconstruction is then realized by minimizing the energy necessary to activate orientation columns among those which are not activated directly by the image. This gives rise to a sub-Riemannian problem on the bundle of directions of the plane.

- Another class of challenging problems concern the functional organization of the motor pathways.

  The interest in establishing a model of the motor pathways, at the same time mathematically rigorous and biologically plausible, comes from the possible spillovers in robotics and neurophysiology. It could help to design better control strategies for robots and artificial limbs, yielding smoother and more progressive movements. Another underlying relevant societal goal (clearly beyond our domain of expertise) is to clarify the mechanisms of certain debilitating troubles such as cerebellar disease, chorea and Parkinson's disease.

  A key issue in order to establish a model of the motor pathways is to determine the criteria underlying the brain's choices. For instance, for the problem of human locomotion (see [14]), identifying such criteria would be crucial to understand the neural pathways implicated in the generation of locomotion trajectories.

  A nowadays widely accepted paradigm is that, among all possible movements, the accomplished ones satisfy suitable optimality criteria (see [76] for a review). One is then led to study an inverse optimal control problem: starting from a database of experimentally recorded movements, identify a cost function such that the corresponding optimal solutions are compatible with the observed behaviors.

  Different methods have been taken into account in the literature to tackle this kind of problems, for instance in the linear quadratic case [46] or for Markov processes [65]. However all these methods have been conceived for very specific systems and they are not suitable in the general case. Two approaches are possible to overcome this difficulty. The direct approach consists in choosing a cost function among a class of functions naturally adapted to the dynamics (such as energy functions) and to compare the solutions of the corresponding optimal control problem to the experimental data. In particular one needs to compute, numerically or analytically, the optimal trajectories and to choose suitable criteria (quantitative and qualitative) for the comparison with observed trajectories. The inverse approach consists in deriving the cost function from the qualitative analysis of the data.

4.3. Switched systems

Switched systems form a subclass of hybrid systems, which themselves constitute a key growth area in automation and communication technologies with a broad range of applications. Existing and emerging areas include automotive and transportation industry, energy management and factory automation. The notion of hybrid systems provides a framework adapted to the description of the heterogeneous aspects related to the interaction of continuous dynamics (physical system) and discrete/logical components.

The characterizing feature of switched systems is the collective aspect of the dynamics. A typical question is that of stability, in which one wants to determine whether a dynamical system whose evolution is influenced by a time-dependent signal is uniformly stable with respect to all signals in a fixed class ([53]).

The theory of finite-dimensional hybrid and switched systems has been the subject of intensive research in the last decade and a large number of diverse and challenging problems such as stabilizability, observability, optimal control and synchronization have been investigated (see for instance [74], [54]).

The question of stability, in particular, because of its relevance for applications, has spurred a rich literature. Important contributions concern the notion of common Lyapunov function: when there exists a Lyapunov function that decays along all possible modes of the system (that is, for every possible constant value of the signal), then the system is uniformly asymptotically stable. Conversely, if the system is stable uniformly with respect to all signals switching in an arbitrary way, then a common Lyapunov function exists [55]. In the *linear* finite-dimensional case, the existence of a common Lyapunov function is actually equivalent to the global uniform exponential stability of the system [61] and, provided that the admissible modes are finitely many, the Lyapunov function can be taken polyhedral or polynomial [21], [22], [34]. A special role in the switched control literature has been played by common quadratic Lyapunov functions, since their existence can be tested rather efficiently (see [35] and references therein). Algebraic approaches to prove the stability of switched systems under arbitrary switching, not relying on Lyapunov techniques, have been proposed in [52], [8].

Other interesting issues concerning the stability of switched systems arise when, instead of considering arbitrary switching, one restricts the class of admissible signals, by imposing, for instance, a dwell time constraint [40].

Another rich area of research concerns discrete-time switched systems, where new intriguing phenomena appear, preventing the algebraic characterization of stability even for small dimensions of the state space [49]. It is known that, in this context, stability cannot be tested on periodic signals alone [24].

Finally, let us mention that little is known about infinite-dimensional switched system, with the exception of some results on uniform asymptotic stability ([58], [71], [72]) and some recent papers on optimal control ([39], [79]).

<p align="center" style="color:red">GEOSTAT Project-Team</p>

4. Application Domains

4.1. Analysis of galactic molecular clouds (GENESIS project)

GENESIS (GENeration and Evolution of Structures in the ISm) is a German/French collaboration project, supported for three years (start 1.5. 2017) by the Deutsche Forschungsgemeinde (DFG) and the Agence national de recherche (ANR). The objective of this research project is to better understand the structure and evolution of molecular clouds in the interstellar medium (ISM) and to link cloud structure with star-formation. For that, far-infrared observations of dust (Herschel) and cooling lines (SOFIA) are combined with ground-based submillimetre observations of molecular lines. Dedicated analysis tools will be used and developed to analyse the maps and compared to simulations in order to disentangle the underlying physical processes such as gravity, turbulence, magnetic fields, and radiation.

4.2. Ocean dynamics, upwelling

Ocean dynamics from remote sensing data is studied in collaboration with LEGOS CNRS Laboratory in Toulouse (SYSCO_2) team, with CRTS and Rabat University in Morocco. The following thematics are studied:

- Large time series analysis on global synoptic ocean data to provide fine characteristics of Ocean turbulence, analysis of altimetry data and ocean/atmosphere exchanges (supported by ICARODE project and submitted LEFE-EC2CO IMECO project).
- Upwelling, Lyapunov exponents and Lagrangian Coherent Structures (supported by TOUBKAL project).

4.3. Non convex optimization for image processing

Three years contract with I2S company on sparse representations and non-convex optimizations methods for image processings problems such as: deconvolution, stiching, noise reduction, HDR mapping etc. This contract is taking place after award-winning results of former GEOSTAT PhD student H. Badri.

4.4. Physiological times series: speech processing

Differential diagnosis between Parkinson's disease and Multiple System Atrophy using complex physiological tile series analysis. Supported by ANR Voice4PD-MSA project.

<p style="text-align:center">**I4S Project-Team**</p>

4. Application Domains

4.1. Civil Engineering

For at least three decades, monitoring the integrity of the civil infrastructure has been an active research topic because of major economical and societal issues, such as durability and safety of infrastructures, buildings and networks. Control of civil structures began a century ago. At stake is the mastering of the ageing of the bridges, as in America (US, Canada) and Great Britain, or the resistance to seismic events and the protection of the cultural heritage, as in Italy and Greece. The research effort in France is very ancient since for example early developments of optical methods to monitor civil structures began in the 70s and SHM practice can be traced back to the 50s with the vibrating wire sensors as strain gauges for dams. Stille the number of sensors actually placed on civil structures is kept to a minimum, mainly for cost reasons, but also because the return on investment sensing and data processing technologies is not properly established for civil structures. One of the current thematic priorities of the C2D2 governmental initiative is devoted to construction monitoring and diagnostics. The picture in Asia (Japan, and also China) is somewhat different, in that recent or currently built bridges are equipped with hundreds if not thousands of sensors, in particular the Hong Kong-Shenzen Western Corridor and Stonecutter Bridge projects. However, the actual use of available data for operational purpose remains unclear.

Among the challenges for vibration-based bridges health monitoring, two major issues are the different kinds of (non measured) excitation sources and the environmental effects. Typically the traffic on *and* under the bridge, the wind and also the rain, contribute to excite the structure, and influence the measured dynamics. Moreover, the temperature is also known to affect the eigenfrequencies and mode-shapes, to an extent which can be significant w.r.t. the deviations to be monitored.

Thermomechanical prestress states affect the dynamic and the static behavior of most bridges, not only of very long and flexible ones. So, the reliable and fast determination of the state of prestress and prestrain associated with a temperature field becomes a crucial step in several engineering processes such as the health monitoring of civil structures. The best possible reconstruction of the temperature field could then become part of a complete process including massively distributed sensing of thermomechanical information on the structure, modeling and algorithms for the on-line detection of damages in the sense of abnormalities with regard to a nominal state, the whole chain being encapsulated in professional tools used by engineers in charge of real-life structural monitoring. For lack of an adequate mobilization of the useful multidisciplinary skills, this way remains about unexplored today.

4.2. Electrical cable and network monitoring

The fast development of electronic devices in modern engineering systems comes with more and more connections through cables, and consequently, the reliability of electric connections becomes a crucial issue. For example, in a modern automotive vehicle, the total length of onboard cables has tremendously increased during the last decades and is now up to 4km. These wires and connectors are subject to ageing or degradation because of severe environmental conditions. In this area, reliability becomes a safety issue. In some other domains, cable defects may have catastrophic consequences. It is thus a crucial challenge to design smart embedded diagnosis systems able to detect wired connection defects in real time. This fact has motivated research projects on methods for fault diagnosis in electric transmission lines and wired networks. Original methods have been recently developed by Inria, notably based on the inverse scattering theory, for cable and network monitoring. Further developments concern both theoretic study and industrial applications.

4. Application Domains

4.1. Energy

In energy, the team mainly focuses on pricing models for demand side management. Demand side management methods are traditionally used to control electricity demand which became quite irregular recently and resulted in inefficiency in supply. We have explored the relationship between energy suppliers and customers who are connected to a smart grid. The smart grid technology allows customers to keep track of hourly prices and shift their demand accordingly, and allows the provider to observe the actual demand response to its pricing strategy. We tackle pricing problems in energy according to the bilevel optimization approaches. Some research works in this domain are supported by bilateral grants with EDF.

4.2. Transportation and Logistics

In transportation and logistics, the team addresses mainly integrated problems, which require taking into account simultaneously different types of decision. Examples are location and routing, inventory management and routing or staff scheduling and warehouse operations management. Such problems occur from the supply chain design level to the logistic facility level. Some research works in this application domain are supported by bilateral grants/contrats with Colisweb, DHL, HappyChic, INFRABEL, or Kéolis.

4.3. Telecommunications

In telecommunications, the team mainly focuses on network design problems and on routing problems. Such problems are optimization problems with complex structure, since the optimization of capacity installation and traffic flow routing have to be addressed simultaneously. Some research works are conducted within a long-term cooperation with Nokia (formerly Alcatel-Lucent Bell Labs).

IPSO Project-Team (section vide)

<p style="text-align:center">**MATHERIALS Project-Team**</p>

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 representation 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 representation 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 $\mathcal{N}_A \sim 6 \times 10^{23}$, 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.

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 10^{18} m^3 of water in the oceans, *i.e.* 7×10^{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 $O(10^{15})$ 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 a few microseconds.

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 representation 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 subjected 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 looking for 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, so that 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 provide 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. The difficulty stems from the fact that the models 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.

MATHRISK Project-Team

4. Application Domains

4.1. Financial Mathematics, Insurance

The applications domains are quantitative finance and insurance with emphasis on risk modeling and control. In particular, Mathrisk focuses on dependence modeling, systemic risk, market microstructure modeling and risk measures.

4. Application Domains

4.1. Aerospace Engineering

Space engineering is very demanding in terms of safe and high-performance control laws. It is therefore prone to fruitful industrial collaborations.

McTAO now has an established expertise in space and celestial mechanics. Our collaborations with industry are mostly on orbit transfer problems with low-thrust propulsion. It can be orbit transfer to put a commercial satellite on station, in which case the dynamics are a Newtonian force field plus perturbations and the small control. There is also, currently, a renewed interest in low-thrust missions such as Lisa Pathfinder (ESA mission towards a Lagrange point of the Sun-Earth system) or BepiColombo (joint ESA-JAXA mission towards Mercury). Such missions look more like a controlled multibody system. In all cases the problem involves long orbit transfers, typically with many revolutions around the primary celestial body. When minimizing time, averaging techniques provide a good approximation. Another important criterion in practice is fuel consumption minimization (crucial because only a finite amount of fuel is onboard a satellite for all its "life"), which amounts to L^1-minimization. Both topics are studied by the team.

Another application is the optimization of space launchers, this lead to issues in stochastic optimization, see section 6.6.1 .

We have a steady relationships with CNES and Thales Alenia Space (Cannes), that have financed or co-financed 3 PhDs and 2 post-docs in the Sophia location of the team in the decade and are a source of inspiration even at the methodological level. Team members also have close connections with Airbus-Safran (Les Mureaux) on launchers.

Some of the authoritative papers in the field were written by team members, with an emphasis on the geometric analysis and on algorithms (coupling of shooting and continuation methods). There are also connections with peers more on the applied side, like D. Scheeres (Colorado Center for Astrodynamics Research at Boulder), the group of F. Bernelli (Politecnico Milano), and colleagues from U. Barcelona (A. Farrès, A. Jorba).

4.2. Magnetic resonance imaging (MRI)

Participants: Bernard Bonnard, Alice Nolot, Jérémy Rouot, Joseph Gergaud, Olivier Cots [ENSEEIHT, Toulouse], Stephen Glaser [TU München, Germany], Dominique Sugny [Univ de Bourgogne].

The starting point of our interest in optimal control for quantum systems was a collaboration with physicist from ICB, University of Burgundy (Dominique Sugny), motivated by an ANR project where we worked on the control of molecular orientation in a dissipative environment using a laser field, and developed optimal control tools, combined with numerical simulations, to analyze the problem for Qubits. This was related to quantum computing rather than MRI.

Using this expertise and under the impulse of Prof. S. Glaser and his group (Chemistry, TU München), we investigated Nuclear Magnetic resonance (NMR) for medical imaging (MRI), where the model is the Bloch equation describing the evolution of the Magnetization vector controlled by a magnetic field, but in fine is a specific Qubit model without decoherence. We worked on, and brought strong contributions to, the contrast problem: typically, given two chemical substances that have an importance in medicine, like oxygenated and de-oxygenated blood, find the (time-dependent) magnetic field that will produce the highest difference in brightness between these two species on the image resulting from Nuclear Magnetic Resonance. This has immediate and important industrial applications in medical imaging. Our contacts are with the above mentioned physics academic labs, who are themselves in contact with major companies.

The team has produced and is producing important work on this problem. One may find a good overview in [46], a reference book has been published on the topic [50], a very complete numerical study comparing different optimization techniques was performed in [44]. We conduct this project in parallel with S. Glaser team, which validated experimentally the pertinence of the methods, the main achievement being the in vivo experiments realized at the Creatis team of Insa Lyon showing the interest to use optimal control methods implemented in modern softwares in MRI in order to produce a better image in a shorter time. A goal is to arrive to a cartography of the optimal contrast with respect to the relaxation parameters using LMI techniques and numerical simulations with the Hamapth and Bocop code; note that the theoretical study is connected to the problem of understanding the behavior of the extremal solutions of a controlled pair of Bloch equations, and this is an ambitious task. Also, one of the difficulties to go from the obtained results, checkable on experiments, to practical control laws for production is to deal with magnetic field space inhomogeneities and poorly known relaxation parameters.

4.3. Swimming at low-Reynolds number

Participants: Bernard Bonnard, Yacine El Alaoui-Faris, Laetitia Giraldi, Clément Moreau, Alice Nolot, Jean-Baptiste Pomet, Jérémy Rouot, Karine Sérier.

Following the historical reference for low Reynolds number locomotion [71], the study of the swimming strategies of micro-organisms is attracting increasing attention in the recent literature. This is both because of the intrinsic biological interest, and for the possible implications these studies may have on the design of bio-inspired artificial replicas reproducing the functionalities of biological systems. In the case of micro-swimmers, the surrounding fluid is dominated by the viscosity effects of the water and becomes reversible. In this regime, it turns out that the infinite dimensional dynamics of the fluid do not have to be retained as state variables, so that the dynamics of a micro-swimmer can be expressed by ordinary differential equations if its shape has a finite number of degrees of freedom. Assuming this finite dimension, and if the control is the rate of deformation, one obtains a control system that is linear (affine without drift) with respect to the controls, *i.e.* the optimal control problem with a quadratic cost defines a sub-Riemannian structure (see section 3.2.3). This is the case where the shape is "fully actuated", *i.e.* if all the variables describing the shape are angles, there is an actuator on each of these angles. For artificial micro-swimmers, this is usually unrealistic, hence (artificial) magneto-elastic micro-swimmers, that are magnetized in order to be deformed by an external magnetic field. In this case, the control functions are the external magnetic field.

In both cases, questions are controllability (straightforward in the fully actuated case), optimal control, possibly path planning. We collaborate with teams that have physical experiments for both.

- In collaboration with D. Takagi and M. Chyba (Univ of Hawaii), this approach is currently at the experimental level for copepod-like swimmer at the university of Hawaii: on the one hand, this zooplankton and its locomotion can be observed, and a robot micro swimmer mimicking a copepod has been constructed, but in fact large enough for direct actuation to be possible, and the low Reynolds number is achieved by using a more viscous fluid. This gives possibilities, through an inverse optimization problem, to determine what cost can be optimised by these crustaceans, see [2], [76], and to validate models on the robot.
- For magneto-elastic micro-robots, Y. El-Alaoui's PhD is co-advised with Stéphane Régnier from the robotics lab ISIR, Univ. Paris 6. Magneto-elastic micro-robots and their magnetic actuation are actually built at ISIR and the aim of the collaboration is to validate models and improve the existing control laws both in performance and in energy; of course, the micro scale does make things difficult.

The questions about optimality of periodic controls raised in section 3.2.5 are related to these applications for periodic deformations, or strokes, playan important role in locomotion.

4.4. Stability of high frequency amplifiers

Participants: Sébastien Fueyo, Jean-Baptiste Pomet, Laurent Baratchart [APICS project-team APICS (FAC-TAS as of 2018)].

Nonlinear hyper-frequency amplifiers contain on the one hand nonlinear active components and on the other hand "lines", that induce some sort of delays and make the system infinite-dimensional. Computer assisted design tools are extensively used. They provide frequency responses but fail to provide a reliable estimation of their stability, and this stability is crucial because an unstable response will not be observed in practice and the engineer needs to have this information between building the actual device. It is, of course, of utmost importance to find means to predict stability/instability from simulations in the Computer assisted design tool.

Sébastien Fueyo's PhD is co-advised between McTAO and APICS on this topic. It is an opportunity to build theoretical basis and justification to a stability analysis through harmonic identification; the latter is one of the specialties of APICS, we collaborate on the infinite-dimensional non-linear stability analysis for periodic solutions and how it works with the results of harmonic identification.

Potential transfer is important.

4.5. Applications of optimal transport

Image processing, biology, fluid mechanics, mathematical physics, game theory, traffic planning, financial mathematics, economics are among the most popular fields of application of the general theory of optimal transport. Many developments have been made in all these fields recently. Three more specific examples:

- In image processing, since a grey-scale image may be viewed as a measure, optimal transportation has been used because it gives a distance between measures corresponding to the optimal cost of moving densities from one to the other, see e.g. the work of J.-M. Morel and co-workers [67].

- In representation and approximation of geometric shapes, say by point-cloud sampling, it is also interesting to associate a measure, rather than just a geometric locus, to a distribution of points (this gives a small importance to exceptional "outlier" mistaken points, see [55]). The relevant distance between measures is again the one coming from optimal transportation.

- A fluid motion or a crowd movement can be seen as the evolution of a density in a given space. If constraints are given on the directions in which these densities can evolve, we are in the framework of non-holonomic transport problems, i.e. these where the cost comes from the point-to-point sub-Riemannian distance, or more general optimal control costs.

<p style="text-align:center">**MEMPHIS Project-Team**</p>

4. Application Domains

4.1. Energy conversion

We consider applications in the domain of wind engineering and sea-wave converters. As an example of application of our methods, we show a recent realization where we model a sea-wave energy converter, see figure 1 . In this unsteady example, the full interaction between the rigid floater, air and water is described by a monolithic model, the Newton's law, where physical parameters such as densities, viscosities and rigidity vary across the domain. The appropriate boundary conditions are imposed at interfaces that arbitrarily cross the grid using adapted schemes built thanks to geometrical information computed via level set functions [37]. The background method for fluid structure interface is the volume penalization method [22] where the level set functions is used to improve the degree of accuracy of the method [26] and also to follow the object. The simulations are unsteady, three dimensional, with $O(10^8)$ grid points on 512 CPUs.
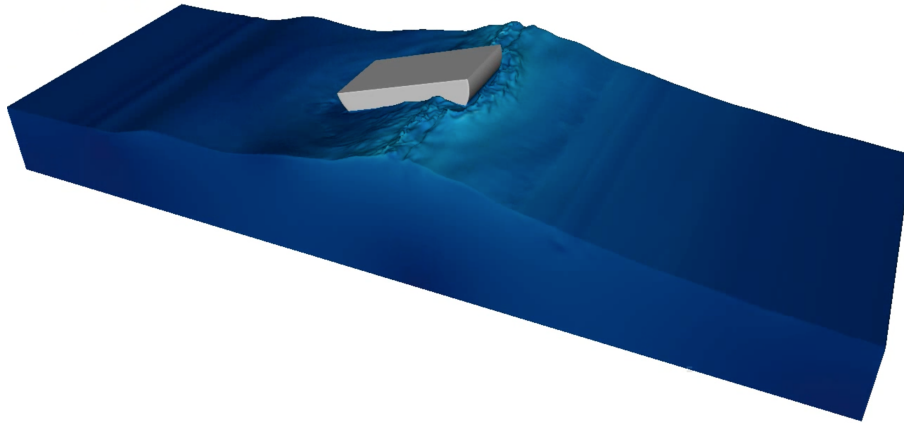
Figure 1. Numerical modeling of a sea-wave converter by a monolithic model and Cartesian meshes.

4.2. Impacts

The numerical modelling of multimaterial rapid dynamics in extreme conditions is an important technological problem for industrial and scientific applications. Experiments are dangerous, need heavy infrastructures and hence are difficult and expensive to realize. The simulation of such phenomena is challenging because they couple large deformations and displacements in solids to strongly non-linear behaviour in fluids. In what follows, we privilege a fully Eulerian approach based on conservation laws, where the different materials are characterized by their specific constitutive laws. This approach was introduced in [30] and subsequently pursued and extended for example in [34], [29], [24], [41].

We study hyper-velocity phenomena where several materials are involved. An example of this approach is the impact of a projectile immersed in air over a shield, see figure 2 . Using the same set of equations across the entire domain, we model the compressible fluid, the hyperelastic material and the interaction at the interface that models possible rebounds. Only the constitutive laws characterize the different materials.

The simulation is performed over a 4000^2 fixed Cartesian grid so that the resulting numerical scheme allows an efficient parallelization (512 processors in this case) with an isomorphism between grid partitioning and processor topology. The challenge for our team is to increase the accuracy of the simulation thanks to grid refinement in the vicinity of the moving interfaces, still guaranteeing scalability and a simple computational set up.
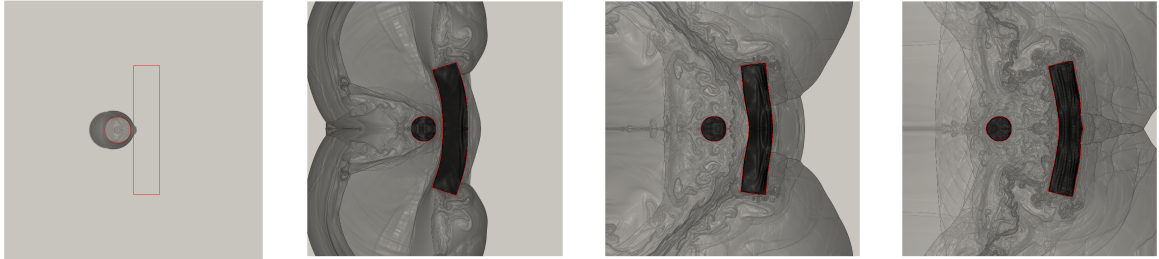
Figure 2. Impact and rebound of a copper projectile on a copper plate. Interface and schlieren at $50\mu s$, $199\mu s$, $398\mu s$ and $710\mu s$. From left to right, top to bottom.

4.3. New materials

Thanks to the multi-scale schemes that we develop, we can characterize new materials from constituents. As an example, consider the material presented in figure 3   left. It is a picture of a dry foam that is used as dielectric material. This micrography is taken at the scale of the dry bubbles, where on the surface of the bubble one can observe the carbon nanotubes as white filaments. The presence of nanotubes in the dry emulsion makes the electrical capacitance of this material significantly affected by its strain state by creating aligned dipoles at a larger scale compared to the size of the dielectric molecules. It is a typical multi-scale phenomenon in presence of widely varying physical properties. This material is used to generate micro currents when it undergoes vibrations. The schemes that we devise allow to model this multi-scale irregular material by a monolithic model (same equation in the whole domain), in this case a variable coefficient diffusion equation. In order to recover adequate accuracy, the numerical scheme is adapted near the interfaces between the different subdomains. The computational hierarchical mesh is directly derived by the micrography of the material (figure 3   right).

4.4. Bio-inspired robotic swimming

In bioinspired robotic swimming the aim is of simulating a three-dimensional swimmer starting from pictures. The first step is to build the three-dimensional fish profile based on two-dimensional data retrieved from the picture of an undeformed fish at rest. This is done by a skeleton technique and a three-dimensional level set function describing the body surface. Then the skeleton is deformed using an appropriate swimming law to obtain a sequence of level set functions corresponding to snapshots of the body surface uniformly taken at different instants.

Thanks to skeleton deformation we typically reconstruct 20% of the snapshots necessary to simulate a swimming stroke, since the time scale of the simulation is significantly smaller than the time step between two subsequent reconstructed snapshots. Also, the surface deformation velocity is required to set the boundary conditions of the flow problem. For this reason it is necessary to build intermediate level set functions and to compute the deformation velocity field between subsequent fish snapshots. Optimal transportation is well suited to achieve this goal providing an objective model to compute intermediate geometries and deformation velocities.
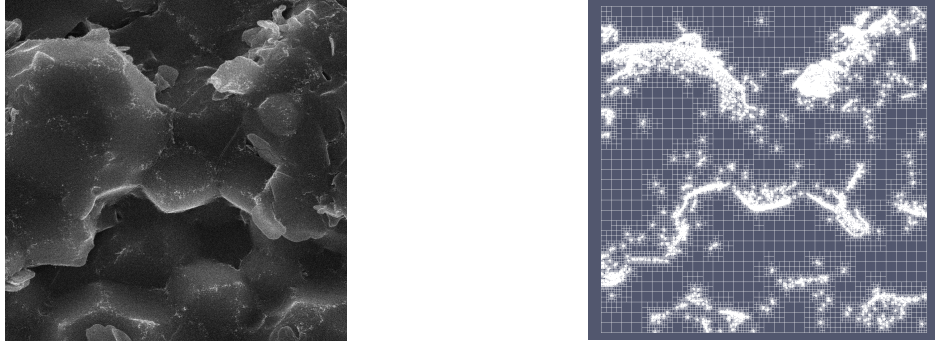
Figure 3. A micrography of an electrostrictive material is shown on the left: the bright regions visualize the carbon nanotubes. The hierarchical grid adapted to the nanotubes is shown on the right. The ratio between the largest and the smallest cell side is 2^7. Project developed in collaboration with the CRPP physics and chemistry lab of the CNRS in Bordeaux (Annie Colin, Philippe Poulin).

Numerical simulations have been performed in 3D, see figure 4 . However, it has been observed that these algorithms do not preserve the physics/features of the represented objects. Indeed, the fish tends to compress during the deformation.
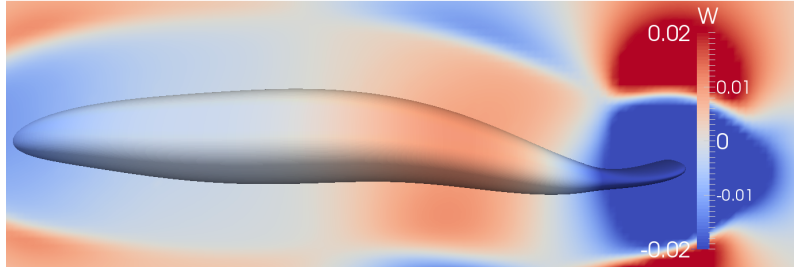
Figure 4. Comparison of the exact deformation velocity (presented inside the swimmer) and the approximated velocity identified using optimal transport (represented outside the fish). The error of the identification scheme is negligible for this component of the velocity, as it can be inferred by comparing the two velocities on the boundary of the swimmer.

For this reason, we will consider incompressible or rigid transports. Another example of bio-inspired swimming is presented in the highlights section.

MEPHYSTO Project-Team (section vide)

MISTIS Project-Team

4. Application Domains

4.1. Image Analysis

Participants: Alexis Arnaud, Aina Frau Pascual, Florence Forbes, Stéphane Girard, Pascal Rubini, Jaime Eduardo Arias Almeida, Pierre-Antoine Rodesch.

As regards applications, several areas of image analysis can be covered using the tools developed in the team. More specifically, in collaboration with team PERCEPTION , we address various issues in computer vision involving Bayesian modelling and probabilistic clustering techniques. Other applications in medical imaging are natural. We work more specifically on MRI and functional MRI data, in collaboration with the Grenoble Institute of Neuroscience (GIN) and the NeuroSpin center of CEA Saclay. We also consider other statistical 2D fields coming from other domains such as remote sensing, in collaboration with Laboratoire de Planétologie de Grenoble. We worked on hyperspectral images. In the context of the "pole de competivité" project I-VP, we worked of images of PC Boards. We also address reconstruction problems in tomography with CEA Grenoble.

4.2. Multi sensor Data Analysis

Participants: Jean-Michel Bécu, Florence Forbes, Thibaud Rahier, Hongliang Lu, Fatima Fofana.

A number of our methods are at the the intersection of data fusion, statistics, machine learning and acoustic signal processing. The context can be the surveillance and monitoring of a zone acoustic state from data acquired at a continuous rate by a set of sensors that are potentially mobile and of different nature (eg WIFUZ project with the ACOEM company in the context of a DGA-rapid initiative). Typical objectives include the development of prototypes for surveillance and monitoring that are able to combine multi sensor data coming from acoustic sensors (microphones and antennas) and optical sensors (infrared cameras) and to distribute the processing to multiple algorithmic blocs. Our interest in acoustic data analysis mainly started from past European projects, POP and Humavips, in collaboration with the PERCEPTION team (PhD theses of Vassil Khalidov, Ramya Narasimha, Antoine Deleforge, Xavier Alameda, and Israel Gebru).

4.3. Biology, Environment and Medicine

Participants: Aina Frau Pascual, Jaime Eduardo Arias Almeida, Alexis Arnaud, Florence Forbes, Stéphane Girard, Emeline Perthame, Jean-Baptiste Durand, Clément Albert, Julyan Arbel, Jean-Michel Bécu, Thibaud Rahier, Brice Olivier, Karina Ashurbekova, Fabien Boux, Veronica Munoz Ramirez.

A third domain of applications concerns biology and medicine. We considered the use of missing data models in epidemiology. We also investigated statistical tools for the analysis of bacterial genomes beyond gene detection. Applications in neurosciences are also considered. In the environmental domain, we considered the modelling of high-impact weather events.

MODAL Project-Team

4. Application Domains

4.1. Economic world

The Modal team applies it research to the economic world through CIFRE Phd supervision such as CA CF (credit scoring), A-Volute (expert in 3D sound), Meilleur Taux (insurance comparator), ... It also has many contracts with companies such as Running care (sport and medical coaching), Arcelor-Mittal (steel industry), Alstom (integrated transport systems) and Vallourec (tubular solutions for the energy markets and for other demanding industrial applications).

4.2. Biology

The second main application domain of the team is the biology. Members of the team are involved in the supervision and scientific animation of the bilille platform, the bioinformatics and bioanalysis platform of Lille. A 2 months research contact has been performed for the Florimond Desprez company (seed industry), and several academic projects have been led in collaboration with research teams in health and biology.

MOKAPLAN Project-Team

4. Application Domains

4.1. Freeform Optics

Following the pioneering work of Caffarelli and Oliker [83], Wang [177] has shown that the inverse problem of freeforming a *convex* reflector which sends a prescribed source to a target intensity is a particular instance of Optimal Transportation. This is a promising approach to automatize the industrial design of optimised energy efficient reflectors (car/public lights for instance). We show in figure 10 the experiment setting and one of the first numerical simulations produced by the ADT Mokabajour.

A third specific topic is the use of optimal transport models in *non-imaging optics*. Light intensity here plays the role of the source/target prescribed mass and the transport map defines the physical shape of specular reflector or refracting lense achieving such a transformation. This models have been around since the works of Oliker and Wang in the 90's. Recent numerical progresses indicate that OT may have an important industrial impact in the design of optical elements and calls for further modelisation and analysis.

The method developed in [58] has been used by researchers of TU Eindhoven in collaboration with Philips Lightning Labs to compute reflectors [159] in a simplified setting (directional light source). Another approach, based on a geometric discretization of Optimal Transportation has been developed in [8], and is able to handle more realistic conditions (punctual light source).

Solving the exact Optimal Transportation model for the Reflector inverse problem involves a generalized Monge-Ampère problem and is linked to the open problem of c-convexity compatible discretization we plan to work on. The corresponding software development is the topic of the ADT Mokabajour.

4.1.1. *Software and industrial output.*

See section 4.3 below for softwares. These methods will clearly become mainstream in reflector design but also in lense design [162]. The industrial problems are mainly on efficiency (light pollution) and security (car head lights) based on free tailoring of the illumination. The figure below is an extreme test case where we exactly reproduce an image. They may represent one of the first incursion on PDE discretisation based methods into the field of non-imaging optics.

4.2. Metric learning for natural language processing

The analysis of large scale datasets to perform un-supervised (clustering) and supervised (classification, regression) learning requires the design of advanced models to capture the geometry of the input data. We believe that optimal transport is a key tool to address this problem because (i) many of these datasets are composed of histograms (social network activity, image signatures, etc.) (ii) optimal transport makes use of a ground metric that enhances the performances of classical learning algorithms, as illustrated for instance in [109].

Some of the theoretical and numerical tools developed by our team, most notably Wasserstein barycenters [41], [67], are now becoming mainstream in machine learning [63], [109]. In its simplest (convex) form where one seeks to only maximize pairwise wasserstein distances, metric learning corresponds to the congestion problem studied by G. Carlier and collaborators [97], [70], and we will elaborate on this connection to perform both theoretical analysis and develop numerical schemes (see for instance our previous work [59]).
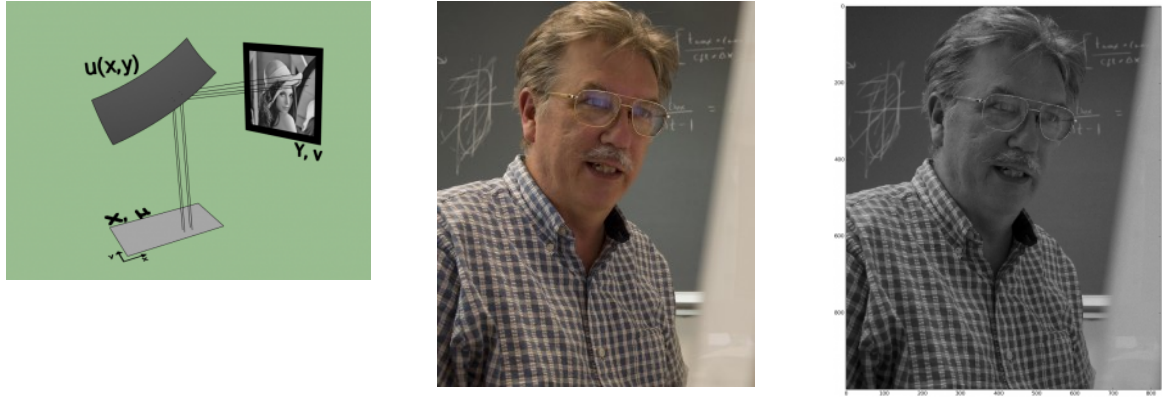
Figure 10. A constant source to a prescribed image (center). The reflector is computed (but not shown) and a resimulation using ray tracing shows the image reflected by the computed reflector.

We aim at developing novel variational estimators extending classification regression energies (SVM, logistic regression [125]) and kernel methods (see [168]). One of the key bottleneck is to design numerical schemes to learn an optimal metric for these purpose, extending the method of Marco Cuturi [108] to large scale and more general estimators. Our main targeted applications is natural language processing. The analysis and processing of large corpus of texts is becoming a key problems at the interface between linguistic and machine learning [45]. Extending classical machine learning methods to this field requires to design suitable metrics over both words and bag-of-words (i.e. histograms). Optimal transport is thus a natural candidate to bring innovative solutions to these problems. In a collaboration with Marco Cuturi (Kyoto University), we aim at unleashing the power of transportation distances by performing ground distance learning on large database of text. This requires to lift previous works on distance on words (see in particular [153]) to distances on bags-of-words using transport and metric learning.

Figure 11. Examples of two histogram (bag-of-words) extracted from the congress speech of US president. In this application, the goal is to infer a meaningful metric on the words of the english language and lift this metric to histogram using OT technics.

4.3. Physics and Astrophysic

The Brenier interpretation of the generalized solutions of Euler equations in the sense of Arnold is an instance of multi-marginal optimal transportation, a recent and expanding research field which also appears in DFT (see chemistry below). Recent numerical developments in OT provide new means of exploring these class of solutions.

In the years 2000 and after the pioneering works of Otto, the theory of *many-particle systems* has become "geometrized" thanks to the observed intimate relation between the geometric theory of geodesic convexity in the Wasserstein distance and the proof of entropy dissipation inequalities that determine the trend to equilibrium. The OT approach to the study of equilibration is still an extremely active field, in particular the various recently established connections to sharp functional inequalities and isoperimetric problems.

4.4. Chemistry

The treatment of *chemical reactions* in the framework of OT is a rather recent development. The classical theory must be extended to deal with the transfer of mass between different particle species by means of chemical reactions.

A promising and significant recent advance is the introduction and analysis of a novel metric that combines the pure transport elements of the Wasserstein distance with the annihilation and creation of mass, which is a first approximation of chemical reactions. The logical next challenge is the extension of OT concepts to vectorial quantities, which allows to rewrite cross-diffusion systems for the concentration of several chemical species as gradient flows in the associated metric. An example of application is the modeling of a *chemical vapor deposition process*, used for the manufacturing of thin-film solar cells for instance. This leads to a degenerate cross-diffusion equations, whose analysis — without the use of OT theory — is delicate. Finding an appropriate OT framework to give the formal gradient flow structure a rigorous meaning would be a significant advance for the applicability of the theory, also in other contexts, like for biological multi-species diffusion.

A very different application of OT in chemistry is a novel approach to the understanding of *density functional theory* (DFT) by using optimal transport with "Coulomb costs", which is highly non convex and singular. Albeit this theory shares some properties with the usual optimal transportation problems, it does not induce a metric between probability measures. It also uses the multi-marginal extension of OT, which is an active field on its own right.

4.5. Biology

OT methods have been introduced in biology via gradient flows in the Wasserstein metric. Writing certain *chemotaxis* systems in variational form allowed to prove sharp estimates on the long time asymptotics of the bacterial aggregation. This application had a surprising payback on the theory: it lead to a better understanding and novel proofs of important functional inequalities, like the logarithmic Hardy-Littlewood-Sobolev inequality. Further applications followed, like transport models for species that avoid over-crowding, or cross-diffusion equations for the description of *biologic segregation*. The inclusion of dissipative cross-diffusion systems into the framework of gradient flows in OT-like metrics appears to be one of the main challenges for the future development of the theory. This extension is not only relevant for biological applications, but is clearly of interest to participants with primary interest in physics or chemistry as well.

Further applications include the connection of OT with game theory, following the idea that many selection processes are based on competition. The ansatz is quite universal and has been used in other areas of the *life sciences* as well, like for the modeling of personal income in economics.

Another application of our methods is the use of inverse problems in measure spaces for microscopy imaging. The *Single Molecule Microscopy Imaging* techniques such as PALM [62] or STORM [164] have yielded a breakthrough in fluorescence microscopy, improving the typical resolution of conventional microscopes (250 nm) by an order of magnitude (20 nm). These techniques convert the problems of full image reconstruction into a family of sparse spike reconstructions. Our variational methods, which take advantage of the sparsity of the signals to recover, are much more powerful than the usual methods used by biologists for sparse recovery. They promise to release the full potential of PALM and STORM in terms of resolution and speed of acquisition.

4.6. Medical Imaging

Applications of variational methods are widespread in medical imaging and especially for diffeomorphic image matching. The formulation of large deformation by diffeomorphisms consists in finding geodesics on a group of diffeomorphisms. This can be seen as a non-convex and smoothed version of optimal transport where a correspondence is sought between objects that can be more general than densities. Whereas the diffeomorphic approach is well established, similarity measures between objects of interest are needed in order to drive the optimization. While being crucial for the final registration results, these similarity measures are often non geometric due to a need of fast computability and gradient computation. However, our team pioneered the use of entropic smoothing for optimal transport which gives fast and differentiable similarity measures that take into account the geometry. Therefore, we expect an important impact on this topic, work still in progress. This example of application belongs to the larger class of inverse problems where a geometric similarity measure such as optimal transport might enhance notably the results. Concerning this particular application, potential interactions with the Inria team ARAMIS and also the team ASCLEPIOS can leverage new proposed similarity measure towards a more applicative impact.

4.7. Economics

Recent years have seen intense cross-fertilization between OT and various problems arising in economics. The principal-agent problem with adverse selection is particularly important in modern microeconomics, mathematically it consists in minimizing a certain integral cost functional among the set of c-concave functions, this problem is convex under some conditions related to the MTW regularity theory for OT as shown in the important paper [116]. Other examples of fruitful interactions between mathematical economics concern multi-marginal OT and multi-populations matching [93], or games with a continuum of agents and Cournot-Nash equilibria [64]. The team has as strong expertise, both numerical and theoretical in the field of variational problems subject to a convexity constraint and their applications to the principal-agent problem. Our expertise in numerical OT and entropic regularization will also enable us to develop efficient solvers for realistic matching and hedonic pricing models.

<p style="text-align:center;color:red;font-weight:bold;">NACHOS Project-Team</p>

4. Application Domains

4.1. Electromagnetic wave propagation

Electromagnetic devices are ubiquitous in present day technology. Indeed, electromagnetism has found and continues to find applications in a wide array of areas, encompassing both industrial and societal purposes. Applications of current interest include (among others) those related to communications (e.g transmission through optical fiber lines), to biomedical devices (e.g microwave imaging, micro-antenna design for telemedicine, etc.), to circuit or magnetic storage design (electromagnetic compatibility, hard disc operation), to geophysical prospecting, and to non-destructive evaluation (e.g crack detection), to name but just a few. Equally notable and motivating are applications in defence which include the design of military hardware with decreased signatures, automatic target recognition (e.g bunkers, mines and buried ordnance, etc.) propagation effects on communication and radar systems, etc. Although the principles of electromagnetics are well understood, their application to practical configurations of current interest, such as those that arise in connection with the examples above, is significantly complicated and far beyond manual calculation in all but the simplest cases. These complications typically arise from the geometrical characteristics of the propagation medium (irregular shapes, geometrical singularities), the physical characteristics of the propagation medium (heterogeneity, physical dispersion and dissipation) and the characteristics of the sources (wires, etc.).

Although many of the above-mentioned application contexts can potentially benefit from numerical modeling studies, the team currently concentrates its efforts on two physical situations.

4.1.1. *Microwave interaction with biological tissues*

Two main reasons motivate our commitment to consider this type of problem for the application of the numerical methodologies developed in the NACHOS project-team:

- First, from the numerical modeling point of view, the interaction between electromagnetic waves and biological tissues exhibit the three sources of complexity identified previously and are thus particularly challenging for pushing one step forward the state-of-the art of numerical methods for computational electromagnetics. The propagation media is strongly heterogeneous and the electromagnetic characteristics of the tissues are frequency dependent. Interfaces between tissues have rather complicated shapes that cannot be accurately discretized using cartesian meshes. Finally, the source of the signal often takes the form of a complicated device (e.g a mobile phone or an antenna array).

- Second, the study of the interaction between electromagnetic waves and living tissues is of interest to several applications of societal relevance such as the assessment of potential adverse effects of electromagnetic fields or the utilization of electromagnetic waves for therapeutic or diagnostic purposes. It is widely recognized nowadays that numerical modeling and computer simulation of electromagnetic wave propagation in biological tissues is a mandatory path for improving the scientific knowledge of the complex physical mechanisms that characterize these applications.

Despite the high complexity both in terms of heterogeneity and geometrical features of tissues, the great majority of numerical studies so far have been conducted using variants of the widely known FDTD method due to Yee [56]. In this method, the whole computational domain is discretized using a structured (cartesian) grid. Due to the possible straightforward implementation of the algorithm and the availability of computational power, FDTD is currently the leading method for numerical assessment of human exposure to electromagnetic waves. However, limitations are still seen, due to the rather difficult departure from the commonly used rectilinear grid and cell size limitations regarding very detailed structures of human tissues. In this context, the general objective of the contributions of the NACHOS project-team is to demonstrate the benefits of high order unstructured mesh based Maxwell solvers for a realistic numerical modeling of the interaction of electromagnetic waves and biological tissues with emphasis on applications related to numerical dosimetry.

Since the creation of the team, our works on this topic have mainly been focussed on the study of the exposure of humans to radiations from mobile phones or wireless communication systems (see Fig. 1). This activity has been conducted in close collaboration with the team of Joe Wiart at Orange Labs/Whist Laboratory http://whist.institut-telecom.fr/en/index.html (formerly, France Telecom Research & Development) in Issy-les-Moulineaux [12].
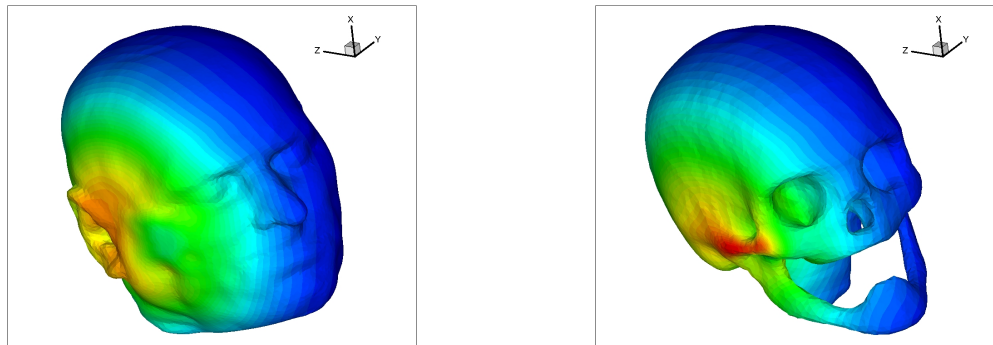
Figure 1. Exposure of head tissues to an electromagnetic wave emitted by a localized source. Top figures: surface triangulations of the skin and the skull. Bottom figures: contour lines of the amplitude of the electric field.

4.1.2. Light/matter interaction on the nanoscale

Nanostructuring of materials has opened up a number of new possibilities for manipulating and enhancing light-matter interactions, thereby improving fundamental device properties. Low-dimensional semiconductors, like quantum dots, enable one to catch the electrons and control the electronic properties of a material, while photonic crystal structures allow to synthesize the electromagnetic properties. These technologies may, e.g., be employed to make smaller and better lasers, sources that generate only one photon at a time, for applications in quantum information technology, or miniature sensors with high sensitivity. The incorporation of metallic structures into the medium add further possibilities for manipulating the propagation of electromagnetic waves. In particular, this allows subwavelength localisation of the electromagnetic field and, by subwavelength structuring of the material, novel effects like negative refraction, e.g. enabling super lenses, may be realized. Nanophotonics is the recently emerged, but already well defined, field of science and technology aimed at establishing and using the peculiar properties of light and light-matter interaction in various nanostructures. Nanophotonics includes all the phenomena that are used in optical sciences for the development of optical devices. Therefore, nanophotonics finds numerous applications such as in optical microscopy, the design of optical switches and electromagnetic chips circuits, transistor filaments, etc. Because of its numerous scientific and technological applications (e.g. in relation to telecommunication, energy production and biomedicine), nanophotonics represents an active field of research increasingly relying on numerical modeling beside experimental studies.

Plasmonics is a related field to nanophotonics. Metallic nanostructures whose optical scattering is dominated by the response of the conduction electrons are considered as plasmomic media. If the structure presents an interface with e.g. a dielectric with a positive permittivity, collective oscillations of surface electrons create surface-plasmons-polaritons (SPPs) that propagate along the interface. SPPs are guided along metal-dielectric interfaces much in the same way light can be guided by an optical fiber, with the unique characteristic of subwavelength-scale confinement perpendicular to the interface. Nanofabricated systems that exploit SPPs offer fascinating opportunities for crafting and controlling the propagation of light in matter. In particular, SPPs can be used to channel light efficiently into nanometer-scale volumes, leading to direct modification of mode dispersion properties (substantially shrinking the wavelength of light and the speed of light pulses

for example), as well as huge field enhancements suitable for enabling strong interactions with non-linear materials. The resulting enhanced sensitivity of light to external parameters (for example, an applied electric field or the dielectric constant of an adsorbed molecular layer) shows great promise for applications in sensing and switching. In particular, very promising applications are foreseen in the medical domain [49]- [57].

Numerical modeling of electromagnetic wave propagation in interaction with metallic nanostructures at optical frequencies requires to solve the system of Maxwell equations coupled to appropriate models of physical dispersion in the metal, such as the Drude and Drude-Lorentz models. Here again, the FDTD method is a widely used approach for solving the resulting system of PDEs [54]. However, for nanophotonic applications, the space and time scales, in addition to the geometrical characteristics of the considered nanostructures (or structured layouts of the latter), are particularly challenging for an accurate and efficient application of the FDTD method. Recently, unstructured mesh based methods have been developed and have demonstrated their potentialities for being considered as viable alternatives to the FDTD method [52]- [53]- [47]. Since the end of 2012, nanophotonics/plasmonics is increasingly becoming a focused application domain in the research activities of the team in close collaboration with physicists from CNRS laboratories, and also with researchers from international institutions.
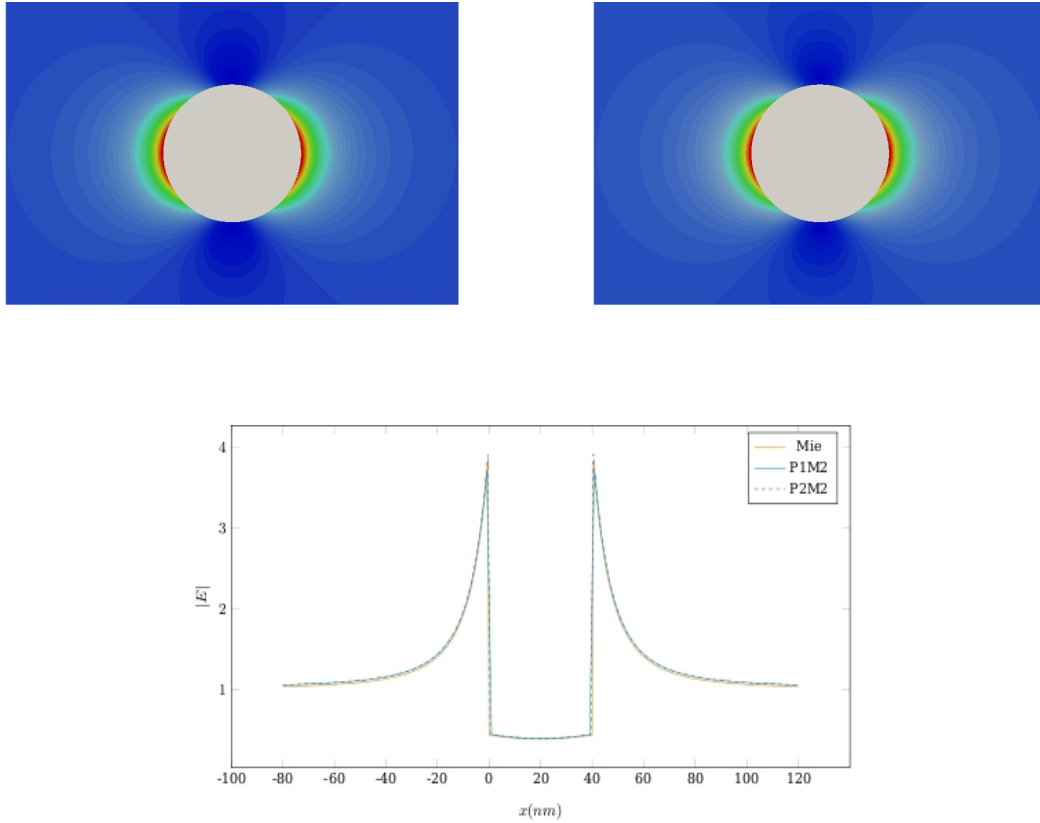
Figure 2. Scattering of a 20 nanometer radius gold nanosphere by a plane wave. The gold properties are described by a Drude dispersion model. Modulus of the electric field in the frequency-domain. Top left figure: Mie solution. Top right figure: numerical solution. Bottom figure: 1d plot of the electric field modulus for various orders of approximation (PhD thesis of Jonathan Viquerat).

4.2. Elastodynamic wave propagation

Elastic wave propagation in interaction with solids are encountered in a lot of scientific and engineering contexts. One typical example is geoseismic wave propagation, in particular in the context of earthquake dynamics or resource prospection.

4.2.1. Earthquake dynamics

To understand the basic science of earthquakes and to help engineers better prepare for such an event, scientists want to identify which regions are likely to experience the most intense shaking, particularly in populated sediment-filled basins. This understanding can be used to improve buildings in high hazard areas and to help engineers design safer structures, potentially saving lives and property. In the absence of deterministic earthquake prediction, forecasting of earthquake ground motion based on simulation of scenarios is one of the most promising tools to mitigate earthquake related hazard. This requires intense modeling that meets the spatial and temporal resolution scales of the continuously increasing density and resolution of the seismic instrumentation, which record dynamic shaking at the surface, as well as of the basin models. Another important issue is to improve the physical understanding of the earthquake rupture processes and seismic wave propagation. Large-scale simulations of earthquake rupture dynamics and wave propagation are currently the only means to investigate these multiscale physics together with data assimilation and inversion. High resolution models are also required to develop and assess fast operational analysis tools for real time seismology and early warning systems.

Numerical methods for the propagation of seismic waves have been studied for many years. Most of existing numerical software rely on finite difference type methods. Among the most popular schemes, one can cite the staggered grid finite difference scheme proposed by Virieux [55] and based on the first order velocity-stress hyperbolic system of elastic waves equations, which is an extension of the scheme derived by Yee [56] for the solution of the Maxwell equations. Many improvements of this method have been proposed, in particular, higher order schemes in space or rotated staggered-grids allowing strong fluctuations of the elastic parameters. Despite these improvements, the use of cartesian grids is a limitation for such numerical methods especially when it is necessary to incorporate surface topography or curved interface. Moreover, in presence of a non planar topography, the free surface condition needs very fine grids (about 60 points by minimal Rayleigh wavelength) to be approximated. In this context, our objective is to develop high order unstructured mesh based methods for the numerical solution of the system of elastodynamic equations for elastic media in a first step, and then to extend these methods to a more accurate treatment of the heterogeneities of the medium or to more complex propagation materials such as viscoelastic media which take into account the intrinsic attenuation. Initially, the team has considered in detail the necessary methodological developments for the large-scale simulation of earthquake dynamics [1]. More recently, the team has initiated a close collaboration with CETE Méditerranée http://www.cete-mediterranee.fr/gb which is a regional technical and engineering centre whose activities are concerned with seismic hazard assessment studies, and IFSTTAR http://www.ifsttar.fr/en/welcome which is the French institute of science and technology for transport, development and networks, conducting research studies on control over aging, risks and nuisances.

4.2.2. Seismic exploration

This application topic is considered in close collaboration with the MAGIQUE-3D project-team at Inria Bordeaux - Sud-Ouest which is coordinating the Depth Imaging Partnership (DIP) http://dip.inria.fr between Inria and TOTAL. The research program of DIP includes different aspects of the modeling and numerical simulation of sesimic wave propagation that must be considered to construct an efficient software suites for producing accurate images of the subsurface. Our common objective with the MAGIQUE-3D project-team is to design high order unstructured mesh based methods for the numerical solution of the system of elastodynamic equations in the time-domain and in the frequency-domain, that will be used as forward modelers in appropriate inversion procedures.
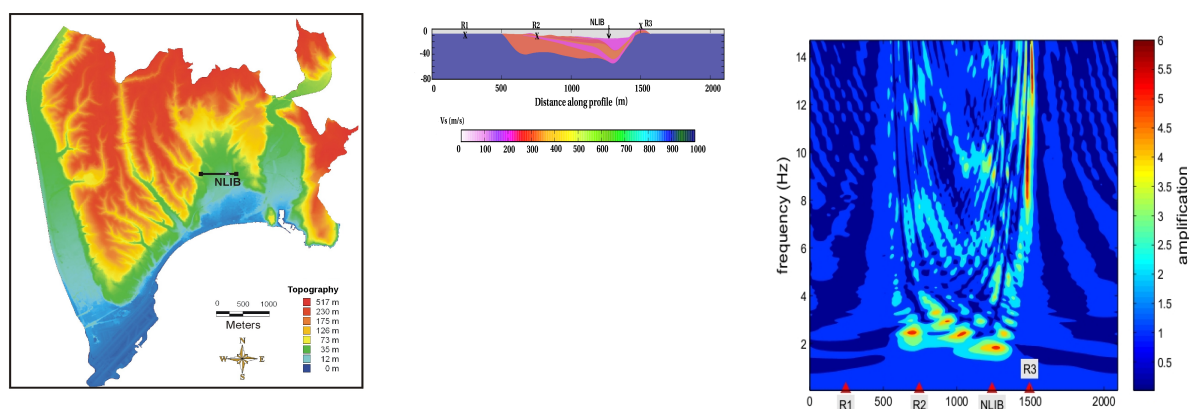
Figure 3. Propagation of a plane wave in a heterogeneous model of Nice area (provided by CETE Méditerranée). Left figure: topography of Nice and location of the cross-section used for numerical simulations (black line). Middle figure: S-wave velocity distribution along the cross-section in the Nice basin. Right figure: transfer functions (amplification) for a vertically incident plane wave ; receivers every 5 m at the surface. This numerical simulation was performed using a numerical method for the solution of the elastodynamics equations coupled to a Generalized Maxwell Body (GMB) model of viscoelasticity (PhD thesis of Fabien Peyrusse).

NANO-D Project-Team (section vide)

NECS Project-Team

4. Application Domains

4.1. A large variety of application domains

Sensor and actuator networks are ubiquitous in modern world, thanks to the advent of cheap small devices endowed with communication and computation capabilities. Potential application domains for research in networked control and in distributed estimation are extremely various, and include the following examples.

- Intelligent buildings, where sensor information on CO_2 concentration, temperature, room occupancy, etc. can be used to control the heating, ventilation and air conditioning (HVAC) system under multi-objective considerations of comfort, air quality, and energy consumption.

- Smart grids: the operation of electrical networks is changing from a centralized optimization framework towards more distributed and adaptive protocols, due to the high number of small local energy producers (e.g., solar panels on house roofs) that now interact with the classic large power-plants.

- Disaster relief operations, where data collected by sensor networks can be used to guide the actions of human operators and/or to operate automated rescue equipment.

- Surveillance using swarms of Unmanned Aerial Vehicles (UAVs), where sensor information (from sensors on the ground and/or on-board) can be used to guide the UAVs to accomplish their mission.

- Environmental monitoring and exploration using self-organized fleets of Autonomous Underwater Vehicles (AUVs), collaborating in order to reach a goal such as finding a pollutant source or tracing a seabed map.

- Infrastructure security and protection using smart camera networks, where the images collected are shared among the cameras and used to control the cameras themselves (pan-tilt-zoom) and ensure tracking of potential threats.

In particular, NECS team is currently focusing in the areas described in detail below.

4.2. Intelligent transportation systems

Throughout the world, roadways are notorious for their congestion, from dense urban network to large freeway systems. This situation tends to get worse over time due to the continuous increase of transportation demand whereas public investments are decreasing and space is lacking to build new infrastructures. The most obvious impact of traffic congestion for citizens is the increase of travel times and fuel consumption. Another critical effect is that infrastructures are not operated at their capacity during congestion, implying that fewer vehicles are served than the amount they were designed for. Using macroscopic fluid-like models, the NECS team has initiated new researches to develop innovative traffic management policies able to improve the infrastructure operations. The research activity is on two main challenges: (1) modeling and forecasting, so as to provide accurate information to users, e.g., travel times; and (2) control, via ramp-metering and/or variable speed limits. The Grenoble Traffic Lab (see http://necs.inrialpes.fr/pages/grenoble-traffic-lab.php) is an experimental platform, collecting traffic infrastructure information in real time from Grenoble South Ring, together with innovative software e.g. for travel-time prediciton, and a show-case where to graphically illustrate results to the end-user. This activity is done in close collaboration with local traffic authorities (DIR-CE, CG38, La Metro), and with the start-up company Karrus (http://www.karrus-its.com/)

4.3. Inertial navigation

Since 2014, the team is exploring techniques for pedestrian navigation and algorithms for attitude estimation, in collaboration with the Tyrex team (Inria-Rhône-Alpes). The goal is to use such algorithms in augmented reality with smartphones. Inertial navigation is a research area related to the determination of 3D attitude and position of a rigid body. Attitude estimation is usually based on data fusion from accelerometers, magnetometers and gyroscopes, sensors that we find usually in smartphones. These algorithms can be used also to provide guidance to pedestrians, e.g., to first responders after a disaster, or to blind people walking in unfamiliar environments. This tasks is particularly challenging for indoor navigation, where no GPS is available.

4.4. Multi-robot collaborative coordination

Due to the cost or the risks of using human operators, many tasks of exploration, or of after-disaster intervention are performed by un-manned drones. When communication becomes difficult, e.g., under water, or in spatial exploration, such robots must be autonomous. Complex tasks, such as exploration, or patrolling, or rescue, cannot be achieved by a single robot, and require a self-coordinated fleet of autonomous devices. NECS team has studied the marine research application, where a fleet of Autonomous Underwater Vehicles (AUVs) self-organize in a formation, adapting to the environment, and reaching a source, e.g., of a pollutant. This has been done in collaboration with IFREMER, within the national project ANR CONNECT and the European FP7 project FeedNetBack. On-going research in the team concerns source localization, with a fleet of mobile robots, including wheeled land vehicles.

4.5. Control design of hydroelectric powerplants

We have started a collaboration with ALSTOM HYDRO, on collaborative and reconfigurable resilient control design of hydroelectric power plants. This work is within the framework of the joint laboratory Inria/ALSTOM (see http://www.inria.fr/innovation/actualites/laboratoire-commun-inria-alstom). A first concrete collaboration has been established with the CIFRE thesis of Simon Gerwig, who has studied how to improve performance of a hydro-electric power-plant outside its design operation conditions, by adaptive cancellation of oscillations that occur in such operation range.

<p style="color:red; text-align:center; font-weight:bold;">NON-A Project-Team</p>

4. Application Domains

4.1. Robots and networked systems

Inria Lille and team FUN are hosting an "equipment of excellence", named FIT-IoT lab. It gives a remote access to thousands of wireless sensors to be connected with hundreds of mobile robots. Today, many sensor scenarios are available, with few robot testbeds.

The package SLIM, developed by Non-A under ROS (Robot Operating System) with the support of an Inria ADT, aims at contributing to this environment. The self deployment of autonomous groups of mobile robots in an unknown and variable environment is a next step for IoT-lab, involving localization, path planning and robust control problems. Our ROS package SLIM aims at combining various algorithms developed by Non-A (localization, path planning, robust control). It should also offer a software library for multi-robot including: optimal local planner based on flatness; plugin for communication between different ROS cores; module Multi-Mapping for robot cooperation; plugin for YEI IMU.

4.2. Living systems: ecological monitoring, modelling, estimation and identification of biological systems, human-computer interaction

Modelling, estimation or detection for living is difficult because such systems cannot be isolated from external influences. Using our numerical differentiation tools, together with modelling techniques, we want to study the following four applications:

- *Biosensing*: Unlike classical approaches deploying physical sensors, biological systems can be used as living sensors. The marine biology lab EPOC (CNRS, Bordeaux) has developed underwater sensors for bivalve molluscs (such as oysters) measuring and sending through RGPS the opening gap between the two valves. We want to use it for water quality monitoring by either identifying oyster's rythm I/O models or by using our differentiation tools. Spawning detection is also considered (ANR WaQMoS).

- *Human-Computer Interaction*: Reduction of the latency between the human input and the system visual response in HCI (ANR TurboTouch). To do that, a simple forecasting algorithm for latency compensation in indirect interaction using a mouse has to be developed based on differentiators.

- *Smart bracelet*: Design a dynamical model for the GSR and for the development of an online algorithm making the GSR signal independent of the user movements. Most resulting computations should be embedded into the bracelet. Collaboration with NEOTROPE (start-up developing a bracelet intended for strong human emotion detection).

- *Microbial populations*: Real-time control of synthetic microbial communities (Inria Project Lab, COSY, under evaluation).

4.3. Turbulent flow control for aircrafts and vehicles

Non-A is active in a Regional consortium gathering micro-technologies (ONERA, IEMN, LAMIH, LML and PPrime lab, Univ. of Poitiers) which aims at developing methods for active control of separated flows (ContrATech subprogram of CPER ELSAT).

Aerodynamic losses are believed to be a major source of energy wastage for a vehicle at speeds higher than 50 km/h. Optimization of the vehicles shapes has reached its limit and such a passive control approach cannot deal with unsteady incoming flow. Similarly, in aeronautics, controlling boundary layer airflow could reduce stall drastically. In such contexts, active control strategies (air blowers, hot film sensors, etc.) are very attractive. But the natural phenomena ruling turbulent flows lead to highly nonlinear and infinite-dimension dynamics. Till now, researchers use either nonlinear PDEs (Navier-Stokes equations) allowing for analysis but improper for control design or unrealistic linear finite-dimension models for classical − but non robust − control. Non-A first wants to propose a model with intermediate complexity (bilinear with time delays, "grey-box" identification on experimental data) and then develop model-based sliding mode and optimal control algorithms.

4.4. Industry and society: i-PID for industry and society, mechatronics (Safran)

- Industry is keen on simple and powerful controllers. The tuning simplicity of the classical PID controller explains its omnipresence in industrial control systems, although its performances drop when the working conditions change. AL.I.E.N SAS was created in 2011 as a spin-off of the Inria project ALIEN, which gave rise to Non-A, working on algebraic estimation and i-PID controller (i.e., using algebraic estimation of the perturbations and apply a simple PID control on some "ultra-local" model). These control technique uses the information contained in the output signal and its estimated derivatives, which can be regarded as "signal-based" controllers. Model-free control technique has been applied in many different domains (electronics, hydroelectric power, etc.).

  Recent research is focused on traffic control and biology. The quality of traffic control laws depends on a good knowledge of the highway characteristics, especially the critical density and the free-flow speed, which are unfortunately most difficult to estimate in real time. Therefore, we aim at developing an algorithm which shows the possibility to control the traffic without the knowledge of density and free-flow speed.

- A collaboration with the Safran Electronics & Defense company has been developed (CIFRE PhD thesis) on the parametric stabilization of gyrostabilized platforms. To do that, we first aim at developing new symbolic-numeric methods for the standard H_∞-loop shaping design problem for models of gyrostabilized platforms in terms of the physical parameters (masses, inertia, etc.) considered as unknown/slowly varying parameters. Using Non-A techniques for the estimation of the physical parameters, we then want to develop new embeddable and adaptive controllers for the robust stabilization of gyrostabilized platforms.

<div align="center">**POEMS Project-Team**</div>

4. Application Domains

4.1. Acoustics

Two particular subjects have retained our attention recently.

1- Aeroacoustics, or more precisely, acoustic propagation in a moving compressible fluid, has been for our team a very challenging topic, which gave rise to a lot of open questions, from the modeling until the numerical approximation of existing models. Our works in this area are partially supported by EADS and Airbus. The final objective is to reduce the noise radiated by Airbus planes.

2- Musical acoustics constitute a particularly attractive application. We are concerned by the simulation of musical instruments whose objectives are both a better understanding of the behavior of existing instruments and an aid for the manufacturing of new instruments. We have successively considered the timpani, the guitar and the piano. This activity is continuing in the framework of the European Project BATWOMAN.

4.2. Electromagnetism

Applied mathematics for electromagnetism during the last ten years have mainly concerned stealth technology and electromagnetic compatibility. These areas are still motivating research in computational sciences (large scale computation) and mathematical modeling (derivation of simplified models for multiscale problems). These topics are developed in collaboration with CEA, DGA and ONERA.

Electromagnetic propagation in non classical media opens a wide and unexplored field of research in applied mathematics. This is the case of wave propagation in photonic crystals, metamaterials or magnetized plasmas.

Finally, the simulation electromagnetic (possibly complex, even fractal) networks is motivated by destructive testing applications. This topic is developed in partnership with CEA-LIST.

4.3. Elastodynamics

Wave propagation in solids is with no doubt, among the three fundamental domains that are acoustics, electromagnetism and elastodynamics, the one that poses the most significant difficulties from mathematical and numerical points of view. A major application topic has emerged during the past years : the non destructive testing by ultra-sounds which is the main topic of our collaboration with CEA-LIST. On the other hand, we are developing efficient integral equation modelling for geophysical applications (soil-structure interaction for civil engineering, seismology).

<div align="center">

QUANTIC Project-Team

</div>

4. Application Domains

4.1. 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-electromechanical 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 border between theoretical and experimental efforts of this emerging field with an emphasis on the applications in quantum information, computation and communication. The main objective of this interdisciplinary team 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 [0] 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 building blocks 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 realizing superconducting circuits at low temperatures and using microwave measurement techniques, 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 quantum information through these integrated circuits.

---

[0]Here the partiality means that no single quantum measurement is capable of providing the complete information on the state of the system.

RANDOPT Team

4. Application Domains

4.1. Applications

Applications of black-box algorithms occur in various domains. Industry but also researchers in other academic domains have therefore a great need to apply black-box algorithms on a daily basis. We see this as a great source of motivation to design better methods. Applications not only allow us to backup our methods and understand what are the relevant features to solve a real-world problem but also help to identify novel difficulties or set priorities in terms of algorithm design.

We are currently dealing with concrete applications related to three industrial collaborations:

- With EDF R&D through the design and placement of bi-facial photovoltaic panels for the postdoc of Asma Atamna funded by the PGMO project NumBER.
- With Thales for the PhD thesis of Konstantinos Varelas (DGA-CIFRE thesis) related to applications in the defense domain.
- With Storengy, a subsidiary of Engie specialized in gas storage, for the PhD thesis of Cheikh Touré.

Another type of application we want to focus on comes from reinforcement learning. The problems addressed in [27] seem to be particularly suited for large-scale variants of CMA-ES.

When dealing with single applications, the results observed are difficult to generalize: typically not many methods are tested on a single application as tests are often time consuming and performed in restrictive settings. Yet, if one circumvent the problem of confidentiality of data and of criticality for companies to publish their applications, real-world problems could become benchmarks as any other analytical function. This would allow to test wider ranges of methods on the problems and to find out whether analytical benchmarks properly capture real-world problem difficulties. We will thus seek to incorporate real-world problems within our COCO platform. This is a recurrent demand by researchers in optimization.

<p style="text-align:center">**RAPSODI Project-Team**</p>

4. Application Domains

4.1. Porous media flows

Porous media flows are of great interest in many contexts, like, e.g., oil engineering, water resource management, nuclear waste repository management, or carbon dioxyde sequestration. We refer to [61], [60] for an extensive discussion on porous media flow models.

From a mathematical point of view, the transport of complex fluids in porous media often leads to possibly degenerate parabolic conservation laws. The porous rocks can be highly heterogeneous and anisotropic. Moreover, the grids on which one intends to solve numerically the problems are prescribed by the geological data, and might be non-conformal with cells of various shapes. Therefore, the schemes used for simulating such complex flows must be particularly robust.

4.2. Corrosion and concrete carbonatation

The team is interested in the theoretical and numerical analysis of mathematical models describing degradation of materials as concrete carbonation and corrosion. The study of such models is an important environmental and industrial issue. Atmospheric carbonation degrades reinforced concretes and limits the lifetime of civil engineering structures. Corrosion phenomena issues occur for instance in the reliability of nuclear power plants and the nuclear waste repository. The study of the long time evolution of these phenomena is of course fundamental in order to predict the lifetime of the structures.

From a mathematical point of view, the modeling of concrete carbonation (see [56]) as the modeling of corrosion in an underground repository (DPCM model developed by Bataillon *et al.* [1]) lead to systems of PDEs posed on moving domains. The coupling between convection-diffusion-reaction equations and moving boundary equations leads to challenging mathematical questions.

4.3. Complex fluid flows

The team is interested in some numerical methods for the simulation of systems of PDEs describing complex flows, like for instance, mixture flows, granular gases, rarefied gases, or quantum fluids.

Let us first focus on fluid mixture flows. The fluid is described by its density, its velocity and its pressure. These quantities obey mass and momentum conservation. On the one hand, when we deal with the 2D variable density incompressible Navier-Stokes equations, we aim to study the ability of the numerical scheme to reproduce some instabilities phenomena such as the Rayleigh-Taylor instability. On the other hand, diffuse interface models have gained renewed interest for the last few years in fluid mechanics applications. From a physical viewpoint, they allow to describe some phase transition phenomena. If the Fick's law relates the divergence of the velocity field to derivatives of the density, one obtains the so called Kazhikhov-Smagulov model [87]. Here, the density of the mixture is naturally highly non homogeneous, and the constitutive law accounts for diffusion effects between the constituents of the mixture. Models of this type can be used for instance to simulate powder-snow avalanches [6], low-Mach flows, or hydrodynamic models arising in combustion theory or transport of pollutants.

Kinetic theory of molecular gases models a gas as a system of elastically colliding spheres, conserving mechanical energy during impact. Once initialized, it takes a molecular gas not more than few collisions per particle to relax to its equilibrium state, characterized by a Maxwellian velocity distribution and a certain homogeneous density (in the absence of external forces). A granular gas is a system of dissipatively colliding, macroscopic particles (grains). This slight change in the microscopic dynamics (converting energy into heat) cause drastic changes in the behavior of the gas: granular gases are open systems, which exhibits self-organized spatio-temporal cluster formations, and has no equilibrium distribution. They can be used to model silos, avalanches, pollen or planetary rings.

The quantum models can be used to describe superfluids, quantum semiconductors, weakly interacting Bose gases or quantum trajectories of Bohmian mechanics. They have attracted considerable attention in the last decades, due in particular to the development of the nanotechnology applications. To describe quantum phenomena, there exists a large variety of models. In particular there exist three different levels of description: microscopic, mesoscopic and macroscopic. The quantum Navier-Stokes equations deal with a macroscopic description in which the quantum effects are taken into account through a third order term called the quantum Bohm potential. This Bohm potential arises from the fluid dynamical formulation of the single-state Schrödinger equation. The non-locality of quantum mechanics is approximated by the fact that the equations of state do not only depend on the particle density but also on its gradient. These equations were employed to model field emissions from metals and steady-state tunneling in metal- insulator- metal structures and to simulate ultra-small semiconductor devices.

4.4. Stratigraphy

The knowledge of the geology is a prerequisite before simulating flows within the subsoil. Numerical simulations of the geological history thanks to stratigraphy numerical codes allow to complete the knowledge of the geology where experimental data are lacking. Stratigraphic models consist in a description of the erosion and sedimentation phenomena at geological scales.

The characteristic time scales for the sediments are much larger than the characteristic time scales for the water in the river. However, the (time-averaged) water flux plays a crucial role in the evolution of the stratigraphy. Therefore, defining appropriate models that take the coupling between the rivers and the sediments into account is fundamental and challenging. Once the models are at hand, efficient numerical methods must be developed.

4.5. Low frequency electromagnetism

Numerical simulation is nowadays an essential tool in order to design electromagnetic systems, by estimating the electromagnetic fields generated in a wide variety of devices. An important challenge for many applications is to quantify the intensity of the electric field induced in a conductor by a current generated in its neighborhood. In the low-frequency regime, we can for example quote the study of the impact on the human body of a high-tension line or, for higher frequencies, the one of a smartphone. But the ability to simulate accurately some electromagnetic fields is also very useful for non destructive control, in the context of the maintenance of nuclear power stations for example. The development of efficient numerical tools, among which the so-called "*a posteriori* error estimators", is consequently necessary to reach a high precision of calculations in order to provide estimations as reliable as possible.

REALOPT Project-Team

4. Application Domains

4.1. Introduction

Our group has tackled applications in logistics, transportation and routing [33] [14] [75], [74], [70], [72], in production planning [92] and inventory control [70], [72], in network design and traffic routing [53], [62], [68], [95], [50], [63], [80], [88], in cutting and placement problems [77], [78], [89], [90], [91], [93], and in scheduling [86], [81], [46], and in High Performance and Cloud Computing [20] [40] [41] [34] [48] [47].

4.2. Network Design and Routing Problems

We are actively working on problems arising in network topology design, implementing a survivability condition of the form "at least two paths link each pair of terminals". We have extended polyhedral approaches to problem variants with bounded length requirements and re-routing restrictions [62]. Associated to network design is the question of traffic routing in the network: one needs to check that the network capacity suffices to carry the demand for traffic. The assignment of traffic also implies the installation of specific hardware at transient or terminal nodes.

To accommodate the increase of traffic in telecommunication networks, today's optical networks use grooming and wavelength division multiplexing technologies. Packing multiple requests together in the same optical stream requires to convert the signal in the electrical domain at each aggregation of disaggregation of traffic at an origin, a destination or a bifurcation node. Traffic grooming and routing decisions along with wavelength assignments must be optimized to reduce opto-electronic system installation cost. We developed and compared several decomposition approaches [97], [96], [95] to deal with backbone optical network with relatively few nodes (around 20) but thousands of requests for which traditional multi-commodity network flow approaches are completely overwhelmed. We also studied the impact of imposing a restriction on the number of optical hops in any request route [94]. We also developed a branch-and-cut approach to a problem that consists in placing sensors on the links of a network for a minimum cost [68], [69].

We studied several time dependent formulations for the unit demand vehicle routing problem [55], [54]. We gave new bounding flow inequalities for a single commodity flow formulation of the problem. We described their impact by projecting them on some other sets of variables, such as variables issued of the Picard and Queyranne formulation or the natural set of design variables. Some inequalities obtained by projection are facet defining for the polytope associated with the problem. We are now running more numerical experiments in order to validate in practice the efficiency of our theoretical results.

We also worked on the p-median problem, applying the matching theory to develop an efficient algorithm in Y-free graphs and to provide a simple polyhedral characterization of the problem and therefore a simple linear formulation [87] simplifying results from Baiou and Barahona.

We considered the multi-commodity transportation problem. Applications of this problem arise in, for example, rail freight service design, "less than truckload" trucking, where goods should be delivered between different locations in a transportation network using various kinds of vehicles of large capacity. A particularity here is that, to be profitable, transportation of goods should be consolidated. This means that goods are not delivered directly from the origin to the destination, but transferred from one vehicle to another in intermediate locations. We proposed an original Mixed Integer Programming formulation for this problem which is suitable for resolution by a Branch-and-Price algorithm and intelligent primal heuristics based on it.

For the problem of routing freight railcars, we proposed two algorithmes based on the column generation approach. These algorithmes have been tested on a set of real-life instances coming from a real Russian freight transportation company. Our algorithms have been faster on these instances than the current solution approach being used by the company.
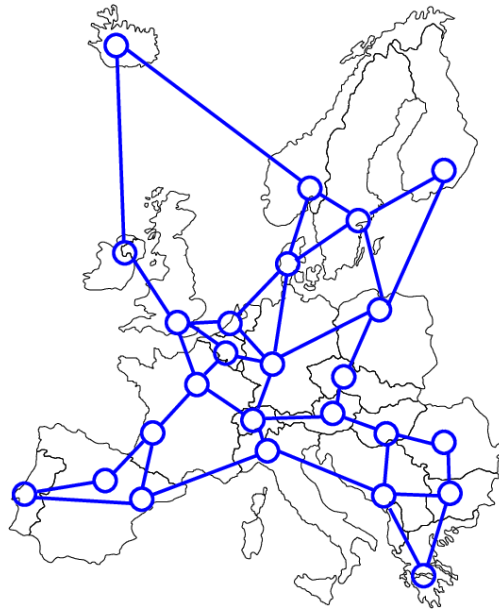
Figure 1. Design of a SDH/SONET european network where demands are multiplexed.

4.3. Packing and Covering Problems

Realopt team has a strong experience on exact methods for cutting and packing problems. These problems occur in logistics (loading trucks), industry (wood or steel cutting), computer science (parallel processor scheduling).

We developed a branch-and-price algorithm for the Bin Packing Problem with Conflicts which improves on other approaches available in the literature [85]. The algorithm uses our methodological advances like the generic branching rule for the branch-and-price and the column based heuristic. One of the ingredients which contributes to the success of our method are fast algorithms we developed for solving the subproblem which is the Knapsack Problem with Conflicts. Two variants of the subproblem have been considered: with interval and arbitrary conflict graphs.

We also developed a branch-and-price algorithm for a variant of the bin-packing problem where the items are fragile. In [44] we studied empirically different branching schemes and different algorithms for solving the subproblems.

We studied a variant of the knapsack problem encountered in inventory routing problem [72]: we faced a multiple-class integer knapsack problem with setups [71] (items are partitioned into classes whose use implies a setup cost and associated capacity consumption). We showed the extent to which classical results for the knapsack problem can be generalized to this variant with setups and we developed a specialized branch-and-bound algorithm.

We studied the orthogonal knapsack problem, with the help of graph theory [65], [64], [67], [66]. Fekete and Schepers proposed to model multi-dimensional orthogonal placement problems by using an efficient representation of all geometrically symmetric solutions by a so called *packing class* involving one *interval graph* for each dimension. Though Fekete & Schepers' framework is very efficient, we have however identified several weaknesses in their algorithms: the most obvious one is that they do not take advantage of the different possibilities to represent interval graphs. We propose to represent these graphs by matrices with consecutive

ones on each row. We proposed a branch-and-bound algorithm for the 2D knapsack problem that uses our 2D packing feasibility check. We are currently developping exact optimization tools for glass-cutting problems in a collaboration with Saint-Gobain [35]. This 2D-3stage-Guillotine cut problems are very hard to solve given the scale of the instance we have to deal with. Moreover one has to issue cutting patterns that avoid the defaults that are present in the glass sheet that are used as raw material. There are extra sequencing constraints regarding the production that make the problem even more complex.

We have also organized a european challenge on packing with society Renault: see http://challenge-esicup-2015.org/. This challenge is about loading trucks under practical constraints.

4.4. Planning, Scheduling, and Logistic Problems

Inventory routing problems combine the optimization of product deliveries (or pickups) with inventory control at customer sites. We considered an industrial application where one must construct the planning of single product pickups over time; each site accumulates stock at a deterministic rate; the stock is emptied on each visit. We have developed a branch-and-price algorithm where periodic plans are generated for vehicles by solving a multiple choice knapsack subproblem, and the global planning of customer visits is coordinated by the master program [73]. We previously developed approximate solutions to a related problem combining vehicle routing and planning over a fixed time horizon (solving instances involving up to 6000 pick-ups and deliveries to plan over a twenty day time horizon with specific requirements on the frequency of visits to customers [75].

Together with our partner company GAPSO from the associate team SAMBA, we worked on the equipment routing task scheduling problem [79] arising during port operations. In this problem, a set of tasks needs to be performed using equipments of different types with the objective to maximize the weighted sum of performed tasks.

We participated to the project on an airborne radar scheduling. For this problem, we developed fast heuristics [61] and exact algorithms [46]. A substantial research has been done on machine scheduling problems. A new compact MIP formulation was proposed for a large class of these problems [45]. An exact decomposition algorithm was developed for the NP-hard maximizing the weighted number of late jobs problem on a single machine [81]. A dominant class of schedules for malleable parallel jobs was discovered in the NP-hard problem to minimize the total weighted completion time [83]. We proved that a special case of the scheduling problem at cross docking terminals to minimize the storage cost is polynomially solvable [84], [82].

Another application area in which we have successfully developed MIP approaches is in the area of tactical production and supply chain planning. In [43], we proposed a simple heuristic for challenging multi-echelon problems that makes effective use of a standard MIP solver. [42] contains a detailed investigation of what makes solving the MIP formulations of such problems challenging; it provides a survey of the known methods for strengthening formulations for these applications, and it also pinpoints the specific substructure that seems to cause the bottleneck in solving these models. Finally, the results of [49] provide demonstrably stronger formulations for some problem classes than any previously proposed. We are now working on planning phytosanitary treatments in vineries.

We have been developing robust optimization models and methods to deal with a number of applications like the above in which uncertainty is involved. In [57], [56], we analyzed fundamental MIP models that incorporate uncertainty and we have exploited the structure of the stochastic formulation of the problems in order to derive algorithms and strong formulations for these and related problems. These results appear to be the first of their kind for structured stochastic MIP models. In addition, we have engaged in successful research to apply concepts such as these to health care logistics [51]. We considered train timetabling problems and their re-optimization after a perturbation in the network [59], [58]. The question of formulation is central. Models of the literature are not satisfactory: continuous time formulations have poor quality due to the presence of discrete decision (re-sequencing or re-routing); arc flow in time-space graph blow-up in size (they can only handle a single line timetabling problem). We have developed a discrete time formulation that strikes a compromise between these two previous models. Based on various time and network aggregation strategies,

we develop a 2-stage approach, solving the contiguous time model having fixed the precedence based on a solution to the discrete time model.

Currently, we are conducting investigations on a real-world planning problem in the domain of energy production, in the context of a collaboration with EDF [26], [27], [28]. The problem consists in scheduling maintenance periods of nuclear power plants as well as production levels of both nuclear and conventional power plants in order to meet a power demand, so as to minimize the total production cost. For this application, we used a Dantzig-Wolfe reformulation which allows us to solve realistic instances of the deterministic version of the problem [60]. In practice, the input data comprises a number of uncertain parameters. We deal with a scenario-based stochastic demand with help of a Benders decomposition method. We are working on Multistage Robust Optimization approaches to take into account other uncertain parameters like the duration of each maintenance period, in a dynamic optimization framework. The main challenge adressed in this work is the joint management of different reformulations and solving techniques coming from the deterministic (Dantzig-Wolfe decomposition, due to the large scale nature of the problem), stochastic (Benders decomposition, due to the number of demand scenarios) and robust (reformulations based on duality and/or column and/or row generation due to maintenance extension scenarios) components of the problem [52].

4.5. Resource Allocation for High Performance and Cloud Computing

In the context of numerical simulations on high performance machines, optimizing data locality and resource usage is very important for faster execution times and lower energy consumption. This optimization can be seen as a special case of scheduling problem on parallel resource, with several challenges. First, instances are typically large: a large matrix factorization (with 50×50 blocks) involves about $30 \cdot 10^3$ tasks. Then, HPC platforms consist of heterogeneous and unrelated resources, what is known to make scheduling problems hard to approximate. Finally, due to co-scheduling effects and shared communication resources, it is not realistic to accurately model the exact duration of tasks. All these observations make it impossible to rely on static optimal solutions, and HPC applications have gone from simple generic static allocations to runtime dynamic scheduling strategies that make their decisions based on the current state of the platform (the location of input data), the expected transfer and running times for the tasks, and some affinity and priority information that have possibly been computed offline. In this context, we are strongly involved in the design of scheduling strategies for the StarPU runtime, with two goals: proving that it is possible to design approximation algorithms whose complexity is extremely small (typically sub-linear in the number of ready tasks), and show that they can be used in practice with good performance results. We are pursuing collaborations both with teams developing the StarPU system (Storm) by designing algorithms for the generic scheduling problems [20], and with teams developing linear algebra algorithms over the runtime (Hiepacs), by proposing specialized algorithms for specific cases. For example, in the case of linear algebra applications on heterogeneous platforms, we have considered the combinatorial optimization problem associated to matrix multiplication, that is amenable to partitioning the unit square into zones of prescribed areas while minimizing the overall size of the boundaries. We have improved the best known approximation ratio to 1.15 in [40] and we have shown that the resulting distribution schemes can indeed be used to design efficient implementations using StarPU in [41].

In the context of Cloud Computing platforms and Data Science, we are interested in resource allocation and data placement strategies. For BigData applications running on data centers platforms, data locality is the largest contributing factor to application performance. In practice, allocation decisions are made at runtime based on strategies that tend to favor local tasks. Our goal is to assess and to improve the efficiency of these runtime strategies. In particular, we have proven that the problem of maximizing locality when allocating map tasks is amenable to a graph orientation and a semi matching problem, what enabled us to assess the efficiency of classical MapReduce allocation algorithm in [34] for the map phase and to propose a low cost algorithm to compute optimal allocation schemes. We also consider more generic VM placement problem for large-scale datacenters. This placement is often handled with naive greedy rules, whereas it is possible to propose online or offline efficient allocation algorithms, for example to optimize the reliability of all applications in a platform with faults [48] or to co-locate applications with compatible periodic load variations [47].

SELECT Project-Team

4. Application Domains

4.1. Introduction

A key goal of SELECT is to produce methodological contributions in statistics. For this reason, the SELECT team works with applications that serve as an important source of interesting practical problems and require innovative methodology to address them. Many of our applications involve contracts with industrial partners, e.g., in reliability, although we also have several academic collaborations, e.g., in genetics and image analysis.

4.2. Curve classification

The field of classification for complex data such as curves, functions, spectra and time series, is an important problem in current research. Standard data analysis questions are being looked into anew, in order to define novel strategies that take the functional nature of such data into account. Functional data analysis addresses a variety of applied problems, including longitudinal studies, analysis of fMRI data, and spectral calibration.

We are focused in particular on unsupervised classification. In addition to standard questions such as the choice of the number of clusters, the norm for measuring the distance between two observations, and vectors for representing clusters, we must also address a major computational problem: the functional nature of the data, which requires new approaches.

4.3. Computer experiments and reliability

For several years now, SELECT has collaborated with the EDF-DER *Maintenance des Risques Industriels* group. One important theme involves the resolution of inverse problems using simulation tools to analyze incertainty in highly complex physical systems.

The other major theme concerns reliability, through a research collaboration with Nexter involving a Cifre convention. This collaboration concerns a lifetime analysis of a vehicle fleet to assess ageing.

Moreover, a collaboration is ongoing with Dassault Aviation on the modal analysis of mechanical structures, which aims to identify the vibration behavior of structures under dynamic excitation. From the algorithmic point of view, modal analysis amounts to estimation in parametric models on the basis of measured excitations and structural response data. In literature and existing implementations, the model selection problem associated with this estimation is currently treated by a rather weighty and heuristic procedure. In the context of our own research, model selection via penalization methods are being tested on this model selection problem.

4.4. Analysis of genomic data

For many years now, SELECT collaborates with Marie-Laure Martin-Magniette (URGV) for the analysis of genomic data. An important theme of this collaboration is using statistically sound model-based clustering methods to discover groups of co-expressed genes from microarray and high-throughput sequencing data. In particular, identifying biological entities that share similar profiles across several treatment conditions, such as co-expressed genes, may help identify groups of genes that are involved in the same biological processes.

Yann Vasseur has completed a thesis co-supervised by Gilles Celeux and Marie-Laure Martin-Magniette on this topic, which is also an interesting investigation domain for the latent block model developed by SELECT. For this work, Yann Vasseur dealt with high-dimensional ill-posed problems where the number of variable was almost equal to the number of observations. He designed heuristic tools using regularized regression methods to circumvent this difficulty.

SELECT collaborates with Anavaj Sakuntabhai and Philippe Dussart (Pasteur Institute) on predicting dengue severity using only low-dimensional clinical data obtained at hospital arrival. An ongoing project also involves statistical meta-analysis of newly collected dengue gene expression data along with recently published data sets from other groups. Further collaborations are underway in dengue fever and encephalitis with researchers at the Pasteur Institute.

SELECT collaborates with Inserm/Paris-Saclay researchers at Kremlin-Bicêtre hospital on cyclic transcriptional clocks and renal corticosteroid signaling, developing statistical tests for synchronous signals.

SELECT is involved in the ANR "jeunes chercheurs" MixStatSeq directed by Cathy Maugis (INSA Toulouse), which is concerned with statistical analysis and clustering of RNASeq genomics data.

4.5. Pharmacovigilance

A collaboration is ongoing with Pascale Tubert-Bitter, Ismael Ahmed and Mohamed Sedki (Pharmacoepidemiology and Infectious Diseases, PhEMI) for the analysis of pharmacovigilance data. In this framework, the goal is to detect, as soon as possible, potential associations between certain drugs and adverse effects, which appeared after the authorized marketing of these drugs. Instead of working on aggregate data (contingency table) like is usually the case, the approach developed aims to deal with individual's data, which perhaps gives more information. Valerie Robert has completed a thesis co-supervised by Gilles Celeux and Christine Keribin on this topic, which involved the development of a new model-based clustering method, inspired by latent block models. Morever, she has defined new tools to estimate and assess the block clustering involved in these models.

4.6. Spectroscopic imaging analysis of ancient materials

Ancient materials, encountered in archaeology and paleontology are often complex, heterogeneous and poorly characterized before physico-chemical analysis. A popular technique to gather as much physico-chemical information as possible, is spectro-microscopy or spectral imaging, where a full spectra, made of more than a thousand samples, is measured for each pixel. The produced data is tensorial with two or three spatial dimensions and one or more spectral dimensions, and requires the combination of an "image" approach with a "curve analysis" approach. Since 2010 SELECT, collaborates with Serge Cohen (IPANEMA) on the development of conditional density estimation through GMM, and non-asymptotic model selection, to perform stochastic segmentation of such tensorial datasets. This technique enables the simultaneous accounting for spatial and spectral information, while producing statistically sound information on morphological and physico-chemical aspects of the studied samples.

SEQUEL Project-Team

4. Application Domains

4.1. Sequential decision making under uncertainty and prediction

The spectrum of applications of our research is very wide: it ranges from the core of our research, that is sequential decision making under uncertainty, to the application of components used to solve this decision making problem.

To be more specific, we work on computational advertizing and recommandation systems; these problems are considered as a sequential matching problem in which resources available in a limited amount have to be matched to meet some users' expectations. The sequential approach we advocate paves the way to better tackle the cold-start problem, and non stationary environments. More generally, these approaches are applied to the optimization of budgeted resources under uncertainty, in a time-varying environment, including constraints on computational times (typically, a decision has to be made in less than 1 ms in a recommandation system). An other field of applications of our research is related to education which we consider as a sequential matching problem between a student, and educational contents.

The algorithms to solve these tasks heavily rely on tools from machine learning, statistics, and optimization. Henceforth, we also apply our work to more classical supervised learning, and prediction tasks, as well as unsupervised learning tasks. The whole range of methods is used, from decision forests, to kernel methods, to deep learning. For instance, we have recently used deep learning on images. We also have a line of works related to software development studying how machine learning can improve the quality of software being developed. More generally, we apply our research to data science.

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: object 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.
- Neuro-imaging: brain-computer interface (fMRI, EEG, MEG).

<p style="text-align:center">**SPHINX Project-Team**</p>

4. Application Domains

4.1. Robotic swimmers

Some companies aim at building biomimetic robots that can swim in an aquarium, as toys (for instance robotswim) but also for medical objectives. The website http://www.robotic-fish.net/ presents a list of several robotic fish that have been built in the last years. Some members of our Inria Project-Team (Munnier, Scheid and Takahashi) developed a collaboration with members of the automatic control laboratory of Nancy CRAN (Daafouz, Jungers) in order to construct a swimming ball in a very viscous fluid. This ball has a macroscopic size but since the fluid is highly viscous, its motion is similar to the motion of a nanorobot. Such nanorobots could be used for medical purposes to carry some medicine or perform small surgical operations. In order to get a better understanding of such robotic swimmers, we have obtained control results via shape changes and we have developed simulation tools ([74], [85]). However, in practice the admissible deformations of the ball are limited since they are realized using piezo-electric actuators. In the future, we want to take these constraints into account by developing two approaches :

1. Solve the control problem by limiting the set of admissible deformations.
2. Find the "best" location of the actuators, in the sense of being the closest to the exact optimal control.

The main tools for this investigation are the 3D codes that we have developed for simulation of fish into a viscous incompressible fluid (SUSHI3D) or into an inviscid incompressible fluid (SOLEIL).

4.2. Aeronautics

We will develop robust and efficient solvers for problems arising in aeronautics (or aerospace) like electromagnetic compatibility and acoustic problems related to noise reduction in an aircraft. Our interest for these issues is motivated by our close contacts with companies like Airbus or Thales Systèmes Aéroportés. We will propose new applications needed by these partners and assist them in integrating these new scientific developments in their home-made solvers. In particular, in collaboration with C. Geuzaine (Université de Liège), we are building a freely available parallel solver based on Domain Decomposition Methods that can handle complex engineering simulations, in terms of geometry, discretization methods as well as physics problems, see http://onelab.info/wiki/GetDDM. Part of this development is done through the grant ANR BECASIM.

4. Application Domains

4.1. Energy Management

Energy management has been one of our priority application fields since 2012, under the lead of Olivier Teytaud. The first works were concerned with sequential decision making, and were based on TAO experience in games, in particular GO, starting with the Associated Team (EA) with Tainan (Taiwan) and the Inria ILAB Metis, in collaborations with SME Artelys. This collaboration continued to be very fruitful, with the ADEME BIA project POST (2014-2017), about long-term investments in power systems, and the ADEME BIA NEXT, that started in April 2017 for 4 years, about the optimization of local grids (at the city or region level). Another line of research is addressed in collaboration with RTE, the company that manages the global French electric network, through Benjamin Donnot's CIFRE PhD.

The collaboration with Artelys had moved from sequential decision making in the Metis ILAB to reinforcement learning, and the design of the Direct Policy Search approach to handle non-anticipativity, in the POST project. Currently, the NEXT project is concerned with the optimization of local networks to meet customer demand, and hightlights the need for an accurate, robust, and fast simulator (Big Data), and some efficient modeling of the demand (Small Data). This is the topic of Victor Berger's PhD (started Oct. 2017). Another issue is directly related to the network optimization - and the optimal setting (possibly online) of graph optimization algorithms, which this is the topic of Herilalaina Rakotoarison's PhD, started Nov. 2017.

The on-going collaboration with RTE is about learning the parries in reaction to network or demand changes to enforce the "n-1" security constraint: at any time, the failure of any of the 30000 links in the network should preserve the security constraints. Logs of network operations over many years are available, but without any "parry" label. This can be achieved by simulating what would have happened without that particular operations regarding the n-1 constraint. The available network simulator is far too slow and sensitive to noise to be useful here. Modeling the network using Deep Networks is straightforward, for a given topology, though computationally costly. The challenge is to take into account the tolopology so that the n-1 constraint can be quickly checked with a single network. The first results on a small grid (118 nodes) outperform the classical DC approximation while providing a significant speedup in calculations [42]. Further works include scaling up, and incorporating all the intricacies of real data.

Several other energy-related works have been, or will be addressed [20], including the organization of a large scale challenge funded by the EU, which was endowed with 2 million euros in prizes (Isabelle Guyon co-organizer), in the context of the EU project See.4C. The participants are asked to predict the power flow on the entire French territory over several years. This challenge will be followed by a challenge in reinforcement learning (RL), in the context of Lisheng Sun's PhD thesis (started Oct. 2016), who is now working on the problem of RL and Automatic Machine Learning (reducing to the largest possible extend human intervention in reinforcement learning). Another direction being explored is the use of causal models to improve explainability of predictive models in decision support systems (Inria-funded post-doc Berna Batu). This should allow us making more intelligible suggestions of corrective actions of operators to bring network operations back to safety when incidents or stress occur.

4.2. Computational Social Sciences: Toward AI Fairness

Several TAU projects are related to computational social and economic sciences. This activity is at the core of the French DataIA *Institut de Convergence*, (head Nozha Boujemaa), gathering 19 partners in the Paris-Saclay area to explore the scientific and ethical impacts of data science and artificial intelligence on the academic, industrial and societal sectors.

Many projects in the domain are related to Causal Modelling (see Section 7.1.1). Some are internal to our team; others involve collaborations with external partners, with a transfer dimension. Others are closely related to some Software platform and are desribed in the corresponding Sections (io.datascience, Section 6.1 and Catolabe, Section 6.3).

- AmiQap (Philippe Caillou, Isabelle Guyon, Michèle Sebag, Paola Tubaro, started 2015). The multivariate analysis of state questionaire data relative to the quality of life at work, in relation with the socio-economical indicators of firms, aims at investigating the relationship between quality of life and economic performances (conditionally to the activity sector), in collaboration with the RITM (U. Paris-Sud), SES (IMTelecom) and La Fabrique de l'Industrie, on data gathered by the Ministry of Labour (DARES). AmiQap is a motivating application for the Causal Modelling studies (PhD Divyan Kalainathan; post-doc Olivier Goudet; coll. David Lopez-Paz, Facebook AI Research).

- Collaborative Hiring (Philippe Caillou, Michèle Sebag, started 2014). Thomas Schmitt's PhD, started in 2014, aims at matching job offers and resumes viewed as a collaborative filtering problem. An alternative approach based on Deep Networks has been developed by François Gonard within his IRT SystemX PhD. The study has been conducted in cooperation with the Web hiring agency Qapa and the non-for-profit organization Bernard Gregory.

- U. Paris-Saclay Nutriperso IRS (Philippe Caillou, Flora Jay, Michèle Sebag, Paola Tubaro) aims to uncover the relationships between health, diets and socio-demographic features. The ultimate goal is to provide personalized *acceptable* recommendations toward healthier eating practices. A milestone is to uncover the causal relationships between diet and health (coll. INRA, INSERM, CEA).

- RESTO (Paola Tubaro, Philippe Caillou). A study of transformations brought about by digital platforms and their effects on the restaurants sector, using a mix of methods that includes both agent-based simulations and machine learning, and fieldwork.

- Sharing Networks (Paola Tubaro, started 2016). Mapping the "collaborative economy" of internet platforms through social network data and analysis.

- OPLa - DiPLab (Paola Tubaro). Two related projects investigating the economy of micro-work platforms in France, and how they integrate with the AI industry ecosystem.

Scientific challenges are related to the FAT (Fairness, Accountability and Transparency) criteria: Metric learning, where the distance/topology to be learned must reflect prior knowledge (e.g. ontologies); Interpretation of clusters built from heterogeneous textual and quantitative data, using the learnt metric/distance; Integration of the human-in-the-loop ("dire d'experts"); Assessment of the models w.r.t. their causality (as opposed to their predictive accuracy) in order to support further interventions.

4.3. High Energy Physics (HEP)

The project started in 2015 with the organization of the Higgs boson ML challenge, in collaboration with the Laboratoire de l'Accelerateur Lineaire (LAL) (David Rousseau and Balazs Kègl) and the ATLAS and CMS projects at CERN. These collaborations have been at the forefront of the broadening interaction between Machine Learning and High Energy Physics, with co-organisation of the Weizmann Hammers and Nails 2017 workshops [44], DataScience@HEP at Fermilab and the Connecting The Dots series.

1. **SystML** (Cécile Germain, Isabelle Guyon, Michèle Sebag, Victor Estrade, Arthur Pesah): Experimental data involve two types of uncertainties: statistical uncertainty (due to natural fluctuations), and systematic uncertainty (due to "known unknowns" such as the imprecise characterization of physics parameters). The SystML project aims to deal with experimental uncertainties along three approaches: i) better calibrating simulators; ii) learning post-processors aimed to filter out the system noise; iii) anticipating the impacts of systematic noise (e.g., on statistical tests) and integrating this impact in the decision process.

V. Estrade's PhD, focusing on the second approach, searches for new data representations insensitive to system-related uncertainty. Taking inspiration from the domain adaptation literature, two strategies have been investigated: i) an agnostric approach based on adversarial supervised learning is used to design an invariant representation (w.r.t. the physics parameters); ii) a prior knowledge-based approach.

2. **TrackML** (Cécile Germain, Isabelle Guyon):
A Tracking Machine Learning challenge (TrackML) [79], [51] is being set up for 1T 2018. Current methods used employed for tracking particles at the LHC (Large Hadron Collider) at CERN will be soon outdated, due to the improved detector apparatus and the associated combinatorial complexity explosion. The LAL and the TAU team have taken a leading role in stimulating both the the ML and HEP communities to renew the toolkit of physicists in preparation for the advent of the next generation of particle detectors.

TrackML refers to recognizing trajectories in the 3D images of proton collisions at the Large Hadron Collider (LHC) at CERN. Think of this as the picture of a fireworks: the time information is lost, but all particle trajectories have roughly the same origin and therefore there is a correspondence between arc length and time ordering. Given the coordinates of the impact of particles on detectors (3D points), the problem is to "connect the dots" or rather the points, i.e. return all sets of points belonging to alleged particle trajectories [16]. From the machine learning point of view, beyond simple clustering, the problem can be treated as a latent variable problem, a tracking problem, or a pattern de-noising problem. A very large dataset (100GB) has been built by the Atlas and CMS collaborations specifically for the challenge.

TrackML will be conducted in 2 phases, the first one favoring innovation over efficiency and the second one aiming at real-time reconstruction. The challenge is supported by Kaggle.

4.4. Autonomous Vehicle

This new application domain builds in fact upon former collaborations of the TAO team with the automotive industry, that created the links with some of the researchers of the R&D departments of Renault (within the Systematic CSDL project and the SystemX ROM project (François Gonard's PhD) and PSA (M. Yagoubi's PhD [84], [85]).

The current work, in collaboration with Renault, is related to the safety of the autonomous vehicle. The validation of the software system is today based on statistics of incidents (failures of some automatized component) assessed from millions of hours of 'driving', either by human drivers in real cars, or by simulations. The work for TAU is related to the set of sample scenarii that are used to compute these statistics. This will require in the first place to identify some latent representation space common to both the actual real-life experiments and the results of the simulation, something that will be achieved using Deep Auto-Encoders of the time series recording the experiments. Two works have started this Fall:

- How to assess the representativity of current set of scenarii, and identify new scenarii to be fed into the simulator to improve the coverage of the scenario space in the common latent representation space, and is the goal of the yet-to-be-signed POC with Renault (Raphaël Jaiswal is working on Renault data since September 2017);

- How to identify original scenarii that lead to failures, an optimization problem in the scenario space. Several criteria for failures will be considered (e.g., getting too close to the preceding car), and the optimization will most likely require building a surrogate model of the simulator for each chosen criterion (and here again Deep Networks are a good candidate), due to its high computing time. This is the topic of Marc Nabhan's CIFRE PhD, started in October 2017 (after a 3 months internship).

4.5. Population Genetics

Work in this application domain started recently, with two main lines of research : dimension reduction of genetic datasets and prediction tasks using genetic data (such as the prediction of past human demography).

- Flora Jay collaborated with Kevin Caye and colleagues (TIMC-IMAG, Grenoble) who developed an R package for inferring coefficients of genetic ancestry, using matrix factorization, alternating quadratic programming and projected least squares algorithms [4]. The extension of ancestry inference and visualization methods to temporal data (for paleogenetics applications) remains to be done.

- The demographic history of one or several population (of any organism) can be partially reconstructed using modern or ancient genetic data. A common approach in the population genetics field is to simulate pseudo-datasets for which the demographic parameters are known and summarize them into handcrafted features. These features are then used as a reference panel in an Approximate Bayesian Computation (likelihood-free) framework. Flora Jay has been developping such methods for the application to whole-genome data [14] [60].

- An untackled challenge in the field is to skip the summary step and directly handle raw data of genetic variations. Théophile Sanchez, who did a 6 month internship in TAU, started his PhD in October 2017 and is currently designing deep learning architectures that are suitable for multi-genome data [33]. In particular these networks should be invariant to the permutation of indiviual genomes and flexible to the input size (see Section 7.2.7).

<p align="center" style="color:red">TOSCA Project-Team</p>

4. Application Domains

4.1. Domain

TOSCA is interested in developing stochastic models and probabilistic numerical methods. Our present motivations come from models with singular coefficients, with applications in Geophysics, Molecular Dynamics and Neurosciences; Lagrangian modeling in Fluid Dynamics and Meteorology; Population Dynamics, Evolution and Genetics; Neurosciences; and Financial Mathematics.

4.1.1. *Stochastic models with singular coefficients: Analysis and simulation*

Stochastic differential equations with discontinuous coefficients arise in Geophysics, Chemistry, Molecular Dynamics, Neurosciences, Oceanography, etc. In particular, they model changes of diffusion of fluids, or diffractions of particles, along interfaces.

For practioners in these fields, Monte Carlo methods are popular as they are easy to interpret — one follows particles — and are in general easy to set up. However, dealing with discontinuities presents many numerical and theoretical challenges. Despite its important applications, ranging from brain imaging to reservoir simulation, very few teams in mathematics worldwide are currently working in this area. The Tosca project-team has tackled related problems for several years providing rigorous approach. Based on stochastic analysis as well as interacting with researchers in other fields, we developed new theoretical and numerical approaches for extreme cases such as Markov processes whose generators are of divergence form with discontinuous diffusion coefficient.

The numerical approximation of singular stochastic processes can be combined with backward stochastic differential equations (BSDEs) or branching diffusions to obtain Monte Carlo methods for quasi-linear PDEs with discontinuous coefficients. The theory of BSDEs has been extensively developed since the 1980s, but the general assumptions for their existence can be quite restrictive. Although the probabilistic interpretation of quasi-linear PDEs with branching diffusions has been known for a long time, there have been only a few works on the related numerical methods.

Another motivation to consider stochastic dynamics in a discontinuous setting came to us from time evolution of fragmentation and coagulation phenomena, with the objective to elaborate stochastic models for the avalanche formation of soils, snow, granular materials or other geomaterials. Most of the models and numerical methods for avalanches are deterministic and involve a wide variety of physical parameters such as the density of the snow, the yield, the friction coefficient, the pressure, the basal topography, etc. One of these methods consists in studying the safety factor (or limit load) problem, related to the shallow flow of a visco-plastic fluid/solid with heterogeneous thickness over complex basal topography. The resulting nonlinear partial differential equation of this last theory involves many singularities, which motivates us to develop an alternative stochastic approach based on our past works on coagulation and fragmentation. Our approach consists in studying the evolution of the size of a typical particle in a particle system which fragments in time.

4.1.2. *Stochastic Lagrangian modeling in Computational Fluid Dynamics*

Stochastic Lagrangian models were introduced in the eighties to simulate complex turbulent flows, particularly two-phase flows. In Computational Fluid Dynamics (CFD), they are intensively used in the so-called Probability Density Functions (PDF) methods in order to model and compute the reaction-phase terms in the fundamental equations of fluid motions. The PDF methods are currently developed in various laboratories by specialists in scientific computation and physicists. However, to our knowledge, we are innovating in two ways:

- our theoretical studies are the pioneering mathematical analysis of Lagrangian stochastic models in CFD;
- our work on the Stochastic Downscaling Method (SDM) for wind simulation is the first attempt to solve the fundamental equations themselves by a fully 3D stochastic particle method.

We emphasize that our numerical analysis is essential to the SDM development which takes benefits from our deep expertise on numerical schemes for McKean-Vlasov-non-linear SDEs.

4.1.3. Population Dynamics, Evolution and Genetics

The activity of the team on stochastic modeling in population dynamics and genetics mainly concerns application in adaptive dynamics, a branch of evolutionary biology studying the interplay between ecology and evolution, ecological modeling, population genetics in growing populations, and stochastic control of population dynamics, with applications to cancer growth modeling. Stochastic modeling in these areas mainly considers individual-based models, where the birth and death of each individual is described. This class of model is well-developed in Biology, but their mathematical analysis is still fragmentary. Another important topic in population dynamics is the study of populations conditioned to non-extinction, and of the corresponding stationary distributions, called quasi-stationary distributions (QSD). This domain has been the object of a lot of studies since the 1960's, but we made recently significant progresses on the questions of existence, convergence and numerical approximation of QSDs using probabilistic tools rather than the usual spectral tools.

Our activity in population dynamics also involves a fully new research project on cancer modeling at the cellular level by means of branching processes. In 2010 the International Society for Protons Dynamics in Cancer was launched in order to create a critical mass of scientists engaged in research activities on Proton Dynamics in Cancer, leading to the facilitation of international collaboration and translation of research to clinical development. Actually, a new branch of research on cancer evolution is developing intensively; it aims in particular to understand the role of proteins acting on cancerous cells' acidity, their effects on glycolysis and hypoxia, and the benefits one can expect from controlling pH regulators in view of proposing new therapies.

4.1.4. Stochastic modeling in Neuroscience

It is generally accepted that many different neural processes that take place in the brain involve noise. Indeed, one typically observes experimentally underlying variability in the spiking times of an individual neuron in response to an unchanging stimulus, while a predictable overall picture emerges if one instead looks at the average spiking time over a whole group of neurons. Sources of noise that are of interest include ionic currents crossing the neural membrane, synaptic noise, and the global effect of the external environment (such as other parts of the brain).

It is likely that these stochastic components play an important role in the function of both the neurons and the networks they form. The characterization of the noise in the brain, its consequences at a functional level and its role at both a microscopic (individual neuron) level and macroscopic level (network of thousands of neurons) is therefore an important step towards understanding the nervous system.

To this end, a large amount of current research in the neuroscientific literature has involved the addition of noise to classical purely deterministic equations resulting in new phenomena being observed. The aim of the project is thus to rigorously study these new equations in order to be able to shed more light on the systems they describe.

4.1.5. Stochastic modeling in Financial Mathematics

4.1.5.1. Technical Analysis

In the financial industry, there are three main approaches to investment: the fundamental approach, where strategies are based on fundamental economic principles; the technical analysis approach, where strategies are based on past price behavior; and the mathematical approach where strategies are based on mathematical models and studies. The main advantage of technical analysis is that it avoids model specification, and thus calibration problems, misspecification risks, etc. On the other hand, technical analysis techniques have limited theoretical justifications, and therefore no one can assert that they are risk-less, or even efficient.

4.1.5.2. Financial Risks Estimation and Hedging

Popular models in financial mathematics usually assume that markets are perfectly liquid. In particular, each trader can buy or sell the amount of assets he/she wants at the same price (the "market price"). They moreover assume that the decision taken by the trader does not affect the price of the asset (the small investor assumption). In practice, the assumption of perfect liquidity is never satisfied but the error due to liquidity is generally negligible with respect to other sources of error such as model error or calibration error, etc.

Derivatives of interest rates are singular for at least two reasons: firstly the underlying (interest rate) is not directly exchangeable, and secondly the liquidity costs usually used to hedge interest rate derivatives have large variation in times.

Due to recurrent crises, the problem of risk estimation is now a crucial issue in finance. Regulations have been enforced (Basel Committee II). Most asset management software products on the markets merely provide basic measures (VaR, Tracking error, volatility) and basic risk explanation features (e.g., "top contributors" to risk, sector analysis, etc).

4.1.5.3. Energy and Carbon Markets

With the rise of renewable energy generation (from solar, wind, waves...), engineers face new challenges which heavily rely on stochastic and statistical problems.

Besides, in the context of the beginning of the second phase (the Kyoto phase) in 2008 of the European carbon market, together with the fact that French carbon tax was scheduled to come into law on Jan. 1, 2010, the year 2009 was a key year for the carbon price modeling. Our research approach adopts the point of view of the legislator and energy producers. We used both financial mathematical tools and a game theory approach. Today, with the third phase of the EU-ETS, that didn't yet start, and the report form the Cour des Comptes (October 2013) that pointed out (among many others point) the lack of mathematical modeling on such carbon market design, we continue our research in this direction.

4.1.5.4. Optimal Stopping Problems

The theory of optimal stopping is concerned with the problem of taking a decision at the best time, in order to maximise an expected reward (or minimise an expected cost). We work on the general problem of optimal stopping with random discounting and additional cost of observation.

4.1.5.5. First hitting times distributions

Diffusion hitting times are of great interest in finance (a typical example is the study of barrier options) and also in Geophysics and Neurosciences. On the one hand, analytic expressions for hitting time densities are well known and studied only in some very particular situations (essentially in Brownian contexts). On the other hand, the study of the approximation of the hitting times for stochastic differential equtions is an active area of research since very few results still are available in the literature.

4. Application Domains

4.1. Discrete event systems (manufacturing systems, networks)

One important class of applications of max-plus algebra comes from discrete event dynamical systems [61]. In particular, modelling timed systems subject to synchronization and concurrency phenomena leads to studying dynamical systems that are non-smooth, but which have remarkable structural properties (nonexpansiveness in certain metrics , monotonicity) or combinatorial properties. Algebraic methods allow one to obtain analytical expressions for performance measures (throughput, waiting time, etc). A recent application, to emergency call centers, can be found in [54].

4.2. Optimal control and games

Optimal control and game theory have numerous well established applications fields: mathematical economy and finance, stock optimization, optimization of networks, decision making, etc. In most of these applications, one needs either to derive analytical or qualitative properties of solutions, or design exact or approximation algorithms adapted to large scale problems.

4.3. Operations Research

We develop, or have developed, several aspects of operations research, including the application of stochastic control to optimal pricing, optimal measurement in networks [108]. Applications of tropical methods arise in particular from discrete optimization [65], [66], scheduling problems with and-or constraints [102], or product mix auctions [116].

4.4. Computing program and dynamical systems invariants

A number of programs and systems verification questions, in which safety considerations are involved, reduce to computing invariant subsets of dynamical systems. This approach appears in various guises in computer science, for instance in static analysis of program by abstract interpretation, along the lines of P. and R. Cousot [69], but also in control (eg, computing safety regions by solving Isaacs PDEs). These invariant sets are often sought in some tractable effective class: ellipsoids, polyhedra, parametric classes of polyhedra with a controlled complexity (the so called "templates" introduced by Sankaranarayanan, Sipma and Manna [109]), shadows of sets represented by linear matrix inequalities, disjunctive constraints represented by tropical polyhedra [55], etc. The computation of invariants boils down to solving large scale fixed point problems. The latter are of the same nature as the ones encountered in the theory of zero-sum games, and so, the techniques developed in the previous research directions (especially methods of monotonicity, nonexpansiveness, discretization of PDEs, etc) apply to the present setting, see e.g. [79], [83] for the application of policy iteration type algorithms, or for the application for fixed point problems over the space of quadratic forms [7]. The problem of computation of invariants is indeed a key issue needing the methods of several fields: convex and nonconvex programming, semidefinite programming and symbolic computation (to handle semialgebraic invariants), nonlinear fixed point theory, approximation theory, tropical methods (to handle disjunctions), and formal proof (to certify numerical invariants or inequalities).

ABS Project-Team (section vide)

AIRSEA Project-Team

4. Application Domains

4.1. The Ocean-Atmosphere System

The evolution of natural systems, in the short, mid, or long term, has extremely important consequences for both the global Earth system and humanity. Forecasting this evolution is thus a major challenge from the scientific, economic, and human viewpoints.

Humanity has to face the problem of **global warming**, brought on by the emission of greenhouse gases from human activities. This warming will probably cause huge changes at global and regional scales, in terms of climate, vegetation and biodiversity, with major consequences for local populations. Research has therefore been conducted over the past 15 to 20 years in an effort to model the Earth's climate and forecast its evolution in the 21st century in response to anthropic action.

With regard to short-term forecasts, the best and oldest example is of course **weather forecasting**. Meteorological services have been providing daily short-term forecasts for several decades which are of crucial importance for numerous human activities.

Numerous other problems can also be mentioned, like **seasonal weather forecasting** (to enable powerful phenomena like an El Niño event or a drought period to be anticipated a few months in advance), **operational oceanography** (short-term forecasts of the evolution of the ocean system to provide services for the fishing industry, ship routing, defense, or the fight against marine pollution) or the prediction of **floods**.

As mentioned previously, mathematical and numerical tools are omnipresent and play a fundamental role in these areas of research. In this context, the vocation of AIRSEA is not to carry out numerical prediction, but to address mathematical issues raised by the development of prediction systems for these application fields, in close collaboration with geophysicists.

AMIBIO Team

4. Application Domains

4.1. Circular RNAs

Participants: Mireille Régnier, Alice Héliou.

Circular RNAs (circRNAs) have been found abundantly in human cells as well as in many other animals. These non-coding RNAs are involved in the regulation of numerous biological processes, and it was recently shown that, as pre-miRNA, they might may actually encode short functional peptides. Our collaborators at Ecole Polytechnique (Biology Dept, LOB) have demonstrated the role of RNA ligase *Pab1020* in RNA circularization. The protein *Pab1020* is a member of the conserved *Rnl3* family of RNA ligases that are predominantly found in hyperthermophiles (archaea, bacteria) and halophiles.

Many computational methods have been proposed to identify and characterize circular RNA from high throughput sequencing data. However, they all suffer from a low specificity, leading to an explosion of false positives. Along with our partners at LOB (Ecole Polytechnique), we develop a robust method for the detection of circRNAs, particularly well-suited to accomodate to analyze sequencing data acquired in extreme environments.

4.2. Analysis of probing data

Participants: Yann Ponty, Mireille Régnier, Afaf Saaidi.

SHAPE probing [47] is an experimental technique in which RNA is exposed to a reagent which, upon reverse-transcription, induces a modification (truncation, mutation) in the DNA. The prevalence of such modifications, which depends on the locally adopted structure(s) (or lack thereof), can be measured for each nucleotide using sequencing techniques, informing regarding the 2D structure. SHAPE probing data can thus be used by structure prediction methods, either to assess their consistency with a proposed structural model, or to restrict the conformation space.

As part of a colloboration with B. Sargueil's lab (Faculté de pharmacie, Paris V) funded by the Fondation pour la Recherche medical, we strive to propose a new paradigm for the analysis data produced using a new experimental technique, called SHAPE analysis (Selective 2'-Hydroxyl Acylation analyzed by Primer Extension). This experimental setup produces an accessibility profile associated with the different positions of an RNA, the *shadow* of an RNA. We currently design new algorithmic strategies to infer the secondary structure of RNA from multiple SHAPE experiments performed by experimentalists at Paris V. Those are obtained on mutants, and will be coupled with a fragment-based 3D modeling strategy developed by our partners at McGill.

4. Application Domains

4.1. Overview

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, bio/geo-chemistry of oceans, resilience of society w.r.t. hazardous flows, urban pollutions, ...

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.

4.2. Geophysical flows

Reduced models like the shallow water equations are particularly well-adapted to the modelling of geophysical flows since there are characterized by large time or/and space scales. For long time simulations, the preservation of equilibria is essential as global solutions are a perturbation around them. The analysis and the numerical preservation of non-trivial equilibria, more precisely when the velocity does not vanish, are still a challenge. In the fields of oceanography and meteorology, the numerical preservation of the so-called geostrophic state, which is the balance between the gravity field and the Coriolis force, can significantly improve the forecasts. In addition, data assimilation is required to improve the simulations and correct the dissipative effect of the numerical scheme.

The sediment transport modelling is of major interest in terms of applications, in particular to estimate the sustainability of facilities with silt or scour, such as canals and bridges. Dredging or filling-up operations are expensive and generally not efficient in the long term. The objective is to determine a configuration almost stable for the facilities. In addition, it is also important to determine the impact of major events like emptying dam which is aimed at evacuating the sediments in the dam reservoir and requires a large discharge. However, the downstream impact should be measured in terms of turbidity, river morphology and flood.

4.3. Hydrological disasters

It is a violent, sudden and destructive flow. Between 1996 and 2005, nearly 80% of natural disasters in the world have meteorological or hydrological origines. The main interest of their study is to predict the areas in which they may occur most probably and to prevent damages by means of suitable amenities. In France, floods are the most recurring natural disasters and produce the worst damages. For example, it can be a cause or a consequence of a dam break. The large surface they cover and the long period they can last require the use of reduced models like the shallow water equations. In urban areas, the flow can be largely impacted by the debris, in particular cars, and this requires fluid/structure interactions be well understood. Moreover, underground flows, in particular in sewers, can accelerate and amplify the flow. To take them into account, the model and the numerical resolution should be able to treat the transition between free surface and underground flows.

Tsunamis are another hydrological disaster largely studied. Even if the propagation of the wave is globally well described by the shallow water model in oceans, it is no longer the case close to the epicenter and in the coastal zone where the bathymetry leads to vertical accretions and produces substantial dispersive effects. The non-hydrostatic terms have to be considered and an efficient numerical resolution should be induced.

While 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).

4.4. Biodiversity and culture

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.

The potential of micro-algae as a source of biofuel and as a technological solution for CO2 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. However, it is clear that many complex physical phenomena have to be added to our model, such as the effect of sunlight on water temperature and density, evaporation and external forcing.

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

4.5. Sustainable energy

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 kinetic and potential parts. The buoys use the potential energy whereas the water-turbines are activated by currents. To become economically relevant, these systems need to be optimized in order to improve their productivity. While for the construction of a harbour, the goal is to minimize swell, in our framework we intend to maximize the wave energy.

This is a complex and original issue which requires a fine model of energy exchanges and efficient numerical tools. In a second step, the optimisation of parameters that can be changed in real-life, such as bottom bathymetry and buoy shape, must be studied. Eventually, physical experiments will be necessary for the validation.

4.6. Urban environment

The urban environment is essentially studied for air and noise pollutions. Air pollution levels and noise pollution levels vary a lot from one street to next. The simulations are therefore carried out at street resolution and take into account the city geometry. The associated numerical models are subject to large uncertainties. Their input parameters, e.g. pollution emissions from road traffic, are also uncertain. Quantifying

the simulation uncertainties is challenging because of the high computational costs of the numerical models. An appealing approach in this context is the use of metamodels, from which ensembles of simulations can be generated for uncertainty quantification.

The simulation uncertainties can be reduced by the assimilation of fixed and mobile sensors. High-quality fixed monitoring sensors are deployed in cities, and an increasing number of mobile sensors are added to the observational networks. Even smartphones can be used as noise sensors and dramatically increase the spatial coverage of the observations. The processing and assimilation of the observations raises many questions regarding the quality of the measurements and the design of the network of sensors.

4.7. SmartCity

There is a growing interest for environmental problems at city scale, where a large part of the population is concentrated and where major pollutions can occur. Numerical simulation is well established to study the urban environment, *e.g.* for road traffic modelling. As part of the smartcity movement, an increasing number of sensors collect measurements, at traditional fixed observation stations, but also on mobile devices, like smartphones. They must properly be taken into account given their number but also their potential low quality.

Pratical applications include air pollution and noise pollution. These directly relate to road traffic. Data assimilation and uncertainty propagation are key topics in these applications.

ARAMIS Project-Team

4. Application Domains

4.1. Introduction

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

- better understanding the pathophysiology of brain disorders;
- designing systems to support clinical decisions such as diagnosis, prognosis and design of clinical trials;
- developing brain computer interfaces for clinical applications.

4.2. Understanding brain disorders

Computational and statistical approaches have the potential to help understand the pathophysiology of brain disorders. We first aim to contribute to better understand the relationships between pathological processes, anatomical and functional alterations, and symptoms. Moreover, within a single disease, there is an important variability between patients. The models that we develop have the potential to identify more homogeneous disease subtypes, that would constitute more adequate targets for new treatments. Finally, we aim to establish the chronology of the different types of alterations. We focus these activities on neurodegeneratives diseases: dementia (Alzheimer's disease, fronto-temporal dementia), Parkinson's disease, multiple sclerosis.

4.3. Supporting clinical decisions

We aim to design computational tools to support clinical decisions, including diagnosis, prognosis and the design of clinical trials. The differential diagnosis of neurodegenerative diseases can be difficult. Our tools have the potential to help clinicians by providing automated classification that can integrate multiple types of data (clinical/cognitive tests, imaging, biomarkers). Predicting the evolution of disease in individual patients is even more difficult. We aim to develop approaches that can predict which alterations and symptoms will occur and when. Finally, new approaches are needed to select participants in clinical trials. Indeed, it is widely recognized that, to have a chance to be successful, treatments should be administered at a very early stage.

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 basis 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 patients' 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.

ASCLEPIOS Project-Team (section vide)

<p style="text-align:center">**ATHENA Project-Team**</p>

4. Application Domains

4.1. Applications of diffusion MRI

Clinical domain: Diagnosis of neurological disorder
Various examples of CNS diseases as Alzheimer's and Parkinson's diseases and others like multiple sclerosis, traumatic brain injury and schizophrenia have characteristic abnormalities in the micro-structure of brain tissues that are not apparent and cannot be revealed reliably by standard imaging techniques. Diffusion MRI can make visible these co-lateral damages to the fibers of the CNS white matter that connect different brain regions. This is why in our research, Diffusion MRI is the structural imaging modality that will be considered to recover the CNS connectivity.

4.2. Applications of M/EEG

Clinical domain: Diagnosis of neurological disorders
The dream of all M/EEG researchers is to alleviate the need for invasive recordings (electrocorticograms or intracerebral electrodes), which are often necessary prior to brain surgery, in order to precisely locate both pathological and vital functional areas. We are involved in this quest, particularly through our collaborations with the La Timone hospital in Marseille.
Subtopics include:

- Diagnosis of neurological disorders such as epilepsy, schizophrenia, tinnitus, ...
- Presurgical planning of brain surgery.

Cognitive research

- Aims at better understanding the brain spatio-temporal organisation.
- Collaboration with the *Laboratory for Neurobiology of Cognition* in order to develop methods that suit their needs for sophisticated data analysis.

Brain Computer Interfaces (BCI) aim to allow direct control of external devices using brain signals such as measured through EEG. In our project, BCI can be seen as an application of EEG processing techniques, but also as an object of fundamental and applied research as they open the way for more dynamical and active brain cognitive protocols.

We are developing research collaborations with the Neurelec company in Sophia Antipolis (subsidiary of Oticon Medical) and with the leading EEG software company BESA based in Munich. We collaborate with Nice University Hospital on the usage of BCI-based communication for ALS [0] patients.

---

[0]Nice University Hospital hosts a regional reference center for patients suffering from Amyotrophic Lateral Sclerosis

BEAGLE Project-Team (section vide)

BIGS Project-Team

4. Application Domains

4.1. Tumor growth-oncology

On this subject, we have new collaborations with clinicians and we want to propose branching processes to model appearance of mutations in tumor. The observed process is the "circulating DNA" (ctDNA). The final purpose is to use ctDNA as a early biomarker of the resistance to an immunotherapy treatment. It is the subject of gthe ITMO project. Another subject is the identification of dynamic network of expression We continue our work on low-grade gliomas. The ongoing collaboration with Montpellier CHU, and a new one with Montreal CRHUM should provide us more data. We initiate as well interactions with researchers from Montreal LIO to extend the previous work. We still have much work to do in modeling to reach our goal of a decision-aid tool for personalised medicine. In the same context, there is a question of clustering analysis of a brain cartography obtained by sensorial simulations during awake surgery.

4.2. Genomic data and micro-organisms population study

Despite of his 'G' in the name of BIGS, Genetics is not central in the applications of the team. However, we want to contribute to a better understanding of the correlations between genes trough their expression data and of the genetic bases of drug response and disease. We have contributed to methods detecting proteomics and transcriptomics variables linked with the outcome of a treatme

4.3. Epidemiology and e-health

We have many works to do in our ongoing projects in the context of personalized medicine with "CHU Nancy". They deal with biomarkers research; prognostic value of quantitative variables and events and scoring, of adverse events. We also want to develop our expertise in rupture detection in a project with APHP for the detection of adverse events, earlier than the clinical signs and symptoms. The clinical relevance of predictive analytics is obvious for high-risk patients such as those with solid organ transplantation or severe chronic respiratory disease for instance. The main challenge is the rupture detection in multivariate and heterogeneous signals (for instance daily measures of electrocardiogram (during 30mn), body temperature, spirometry parameters, sleep duration, etc ... Other collaborations with clinicians concern foetopathology and we want to use our work on conditional distribution function to explain fetal and child growth. We have data from the "Service de foetopathologie et de placentologie" of the "Maternité Régionale Universitaire" (CHU Nancy).

4.4. Dynamics of telomeres

The telomeres are disposable buffers at the ends of chromosomes which are truncated during cell division; so that, over time, due to each cell division, the telomere ends become shorter. By this way, they are markers of aging. Trough a beginning collaboration with Pr A. Benetos, geriatrician at CHU Nancy, we recently data on the distribution of the length of telomeres from blood cells. With some members of Inria team TOSCA, we want to work in three connected directions: (1) refine methodology for the analysis of the available data; (2) propose a dynamical model for the lengths of telomeres and study its mathematical properties (long term behavior, quasi-stationarity, etc); and (3) use these properties to develop new statistical methods. A slot of postdoc position is already planned in the Lorraine Université d'Excellence, LUE project GEENAGE (managed by CHU Nancy).

<p style="text-align:center; color:red;">**BIOCORE Project-Team**</p>

4. Application Domains

4.1. Bioenergy

Finding sources of renewable energy is a key challenge for our society. We contribute to this topic through two main domains for which a strong and acknowledged expertise has been acquired over the years. First, we consider anaerobic digesters, the field of expertise of the members of the team at the Laboratory of Enviromental Biotechnology (LBE), for the production of methane and/or biohydrogen from organic wastes. The main difficulty is to make these processes more reliable and exploit more efficiently the produced biogas by regulating both its quality and quantity despite high variability in the influent wastes. One of the specific applications that needs to be tackled is the production of biogas in a plant when the incoming organic waste results from the mixing of a finite number of substrates. The development of control laws that optimize the input mix of the substrates as a function of the actual state of the system is a key challenge for the viability of this industry.

The second topic consists in growing microalgae, the field of expertise of the members of the team at the Oceanographic Laboratory of Villefranche-sur-Mer (LOV), to produce biofuel. These microorganisms can synthesize lipids with a much higher productivity than terrestrial oleaginous species. The difficulty is to better understand the involved processes, which are mainly transient, to stimulate and optimize them on the basis of modeling and control strategies. Predicting and optimizing the productivity reached by these promising systems in conditions where light received by each cell is strongly related to hydrodynamics, is a crucial challenge.

Finally, for the energy balance of the process, it is important to couple microalgae and anaerobic digestion to optimize the solar energy that can be recovered from microalgae, as was explored within the ANR Symbiose project (2009-2012) [3].

4.2. CO_2 fixation and fluxes

Phytoplanktonic species, which assimilate CO_2 during photosynthesis, have received a lot of attention in the last years. Microalgal based processes have been developed in order to mitigate industrial CO_2. As for biofuel productions, many problems arise when dealing with microalgae which are more complex than bacteria or yeasts. Several models have been developed within our team to predict the CO_2 uptake in conditions of variable light and nitrogen availability. The first modeling challenge in that context consists in taking temperature effects and light gradient into account.

The second challenge consists in exploiting the microalgal bioreactors which have been developed in the framework of the quantification of carbon fluxes between ocean and atmospheres. The SEMPO platform (simulator of variable environment computer controlled), developed within the LOV team, has been designed to reproduce natural conditions that can take place in the sea and to accurately measure the cells behavior. This platform, for which our team has developed models and control methods over the years, is an original and unique tool to develop relevant models which stay valid in dynamic conditions. It is worth noting that a better knowledge of the photosynthetic mechanisms and improved photosynthesis models will benefit both thematics: CO_2 mitigation and carbon fluxes predictions in the sea.

4.3. Biological control for plants and micro-plants production systems

This research concentrates on the protection of cultures of photosynthetic organisms against their pests or their competitors. The cultures we study are crop and micro-algae productions. In both cases, the devices are more or less open to the outside, depending on the application (greenhouse/field, photobioreactor/raceway), so that they may give access to harmful pathogens and invading species. We opt for protecting the culture through the use of biocontrol in a broad sense.

In crop production, biocontrol is indeed a very promising alternative to reduce pesticide use: it helps protecting the environment, as well as the health of consumers and producers; it limits the development of resistance (compared to chemicals)... The use of biocontrol agents, which are, generically, natural enemies (predators, parasitoids or pathogens) of crop pests [66], is however not widespread yet because it often lacks efficiency in real-life crop production systems (while its efficiency in the laboratory is much higher) and can fail to be economically competitive. Resistant crops are also used instead of pesticides to control pests and pathogens, but the latter eventually more or less rapidly overcome the resistance, so these crops need to be replaced by new resistant crops. As resistant genes are a potentially limited resource, a challenge is to ensure the durability of crop resistance. Our objective is to propose models that would help to explain which factors are locks that prevent the smooth transition from the laboratory to the agricultural crop, as well as develop new methods for the optimal deployment of the pests natural enemies and of crop resistance.

Microalgae production is faced with exactly the same problems since predators of the produced microalgae (e.g. zooplankton) or simply other species of microalgae can invade the photobioreactors and outcompete or eradicate the one that we wish to produce. Methods need therefore to be proposed for fighting the invading species; this could be done by introducing predators of the pest and so keeping it under control, or by controling the conditions of culture in order to reduce the possibility of invasion; the design of such methods could greatly take advantage of our knowledge developed in crop protection since the problems and models are related.

4.4. Biological depollution

These works will be carried out with the LBE , mainly on anaerobic treatment plants. This process, despite its strong advantages (methane production and reduced sludge production) can have several locally stable equilibria. In this sense, proposing reliable strategies to stabilize and optimise this process is a key issue. Because of the recent (re)development of anaerobic digestion, it is crucial to propose validated supervision algorithms for this technology. A problem of growing importance is to take benefit of various waste sources in order to adapt the substrate quality to the bacterial biomass activity and finally optimize the process. This generates new research topics for designing strategies to manage the fluxes of the various substrate sources meeting at the same time the depollution norms and providing a biogas of constant quality. In the past years, we have developed models of increasing complexity. However there is a key step that must be considered in the future: how to integrate the knowledge of the metabolisms in such models which represent the evolution of several hundreds bacterial species? How to improve the models integrating this two dimensional levels of complexity? With this perspective, we wish to better represent the competition between the bacterial species, and drive this competition in order to maintain, in the process, the species with the highest depollution capability. This approach, initiated in [71] must be extended from a theoretical point of view and validated experimentally.

BIOVISION Team

4. Application Domains

4.1. Applications of cross reality technologies for low-vision

- High tech vision aid systems for low-vision patients: Vision aid systems for low-vision patients is an application domain with commercial products already existing. A variety of XR solutions are on the market using different kinds of platforms (dedicated or large public ones) and offering different kinds of functionalities (magnification, enhancement, text to speech, face and object recognition). Our goal is to propose new solutions to help patients in a given task.

- Serious games [0] for rehabilitation: Since cross reality technology is now becoming available for the large public, it is a promising platform to develop new rehabilitation exercises.

- Cognitive research: A new trend in cross reality technology is to include eye-tracking so that these platforms could also be very efficient to conduct cognitive and behavioural research on a large scale.

4.2. Applications of vision modeling studies

- Neuroscience research: Making in-silico experiment is a way to reduce the experimental costs, to test hypotheses and design models, and to test algorithms. Our goal is to develop a large-scale simulations platform of impaired retinas, allowing to mimic specific degeneracies or pharmacologically induced impairments, as well as to emulate electric stimulation by prostheses. In addition, the plateform provides a realistic entry to models or simulators of the thalamus or the visual cortex, in contrast to the entries usually considered.

- Education: The simulation plateform we develop could also be a useful tool for educational purposes, illustrating for students how the retina works and respond to visual stimuli.

---

[0]Serious games are games designed for a primary purpose which is not pure entertainment. In our context, we think about serious games as a way to help low vision patients in performing rehabilitation exercises.

BONSAI Project-Team

4. Application Domains

4.1. Life Sciences and health

Our research plays a pivotal role in all fields of life sciences and health where genomic data are involved. This includes more specifically the following topics: plant genomics (genome structure, evolution, microRNAs), cancer (leukemia, mosaic tumors), drug design (NRPSs), environment (metagenomics and metatranscriptomics), virology (evolution, RNA structures) ...

CAMIN Team (section vide)

<p style="text-align:center;color:red;">CAPSID Project-Team</p>

4. Application Domains

4.1. Biomedical Knowledge Discovery

Participants: Marie-Dominique Devignes [contact person], David Ritchie, Sabeur Aridhi, Gabin Personeni, Seyed Ziaeddin Alborzi, Bishnu Sarker, Claire Lacomblez.

This projects in this domain are carried out in collaboration with the Orpailleur Team.

Huge and ever increasing amounts of biomedical data ("Big Data") are bringing new challenges and novel opportunities for knowledge discovery in biomedicine. We are actively collaborating with biologists and clinicians to design and implement approaches for selecting, integrating, and mining biomedical data in various areas. In particular, we are focusing on leveraging bio-ontologies at all steps of this process (the main thesis topic of Gabin Personeni, co-supervised by Marie-Dominique Devignes and Adrien Coulet from the Orpailleur team). One specific application concerns exploiting Linked Open Data (LOD) to characterise the genes responsible for intellectual deficiency. This work is in collaboration with Pr. P. Jonveaux of the Laboratoire de Génétique Humaine at CHRU Nancy [54], [55]. This involves using inductive logic programming as a machine learning method and at least three different ontologies (Gene Ontology, Human Phenotype Ontology, and Disease Ontology). This approach has also been applied using pattern structure mining (an extension of formal concept analysis) of drug and disease ontologies to discover frequently associated adverse drug events in patients [20]. This work was performed in collaboration with the Centre for BioMedical Informatics Research (BMIR) at Stanford University.

Recently, a new application for biomedical knowledge discovery has emerged from the ANR "FIGHT-HF" (fight heart failure) project, which is in collaboration with several INSERM teams at CHRU Nancy. In this case, the molecular mechanisms that underly HF at the cellular and tissue levels will be considered against a background of all available data and ontologies, and represented in a single integrated complex network. A network platform is under construction with the help of a young start-up company called Edgeleap. Together with this company, we are developing query and analysis facilities to help biologists and clinicians to identify relevant biomarkers for patient phenotyping [34]. Docking of small molecules on candidate receptors, as well as protein-protein docking will also be used to clarify a certain number of relations in the complex HF network.

4.2. Prokaryotic Type IV Secretion Systems

Participants: Marie-Dominique Devignes [contact person], Bernard Maigret, Isaure Chauvot de Beauchêne, David Ritchie.

Prokaryotic type IV secretion systems constitute a fascinating example of a family of nanomachines capable of translocating DNA and protein molecules through the cell membrane from one cell to another [26]. The complete system involves at least 12 proteins. The structure of the core channel involving three of these proteins has recently been determined by cryo-EM experiments [40], [63]. However, the detailed nature of the interactions between the remaining components and those of the core channel remains to be resolved. Therefore, these secretion systems represent another family of complex biological systems (scales 2 and 3) that call for integrated modeling approaches to fully understand their machinery.

In the frame of the Lorraine Université d'Excellence (LUE) "CITRAM" project MD Devignes is pursuing her collaboration with Nathalie Leblond of the Genome Dynamics and Microbial Adaptation (DynAMic) laboratory (UMR 1128, Université de Lorraine, INRA) on the discovery of new integrative conjugative elements (ICEs) and integrative mobilisable elements (IMEs) in prokaryotic genomes. These elements use Type IV secretion systems for transferring DNA horizontally from one cell to another. We have discovered more than 200 new ICEs/IMEs by systematic exploration of 72 Streptococcus genome. As these elements encode all or a subset of the components of the Type IV secretion system, they constitute a valuable source of

sequence data and constraints for modeling these systems in 3D. Another interesting aspect of this particular system is that unlike other secretion systems, the Type IV secretion systems are not restricted to a particular group of bacteria [14].

4.3. Protein-RNA Interactions

Participants: Isaure Chauvot de Beauchêne [contact person], Bernard Maigret, Maria Elisa Ruiz Echartea, David Ritchie.

As well as playing an essential role in the translation of DNA into proteins, RNA molecules carry out many other essential biological functions in cells, often through their interactions with proteins. A critical challenge in modelling such interactions computationally is that the RNA is often highly flexible, especially in single-stranded (ssRNA) regions of its structure. These flexible regions are often very important because it is through their flexibility that the RNA can adjust its 3D conformation in order to bind to a protein surface. However, conventional protein-protein docking algorithms generally assume that the 3D structures to be docked are rigid, and so are not suitable for modeling protein-RNA interactions. There is therefore much interest in developing protein-RNA docking algorithms which can take RNA flexibility into account.

We are currently developing a novel flexible docking algorithm which first docks small fragments of ssRNA (typically three nucleotides at a time) onto a protein surface, and then combinatorially reassembles those fragments in order to recover a contiguous ssRNA structure on the protein surface [33], [32]. We have since implemented a prototype "forward-backward" dynamic programming algorithm with stochastic backtracking that allows us to model protein RNA interactions for ssRNAs of up to 7 nucleotides without requiring any prior knowledge of the interaction, while still avoiding a brute-force search. In the frame of our PEPS collaboration "InterANRIL" with the IMoPA lab (Univ de Lorraine), we are currently working with biologists to apply the approach to modeling certain long non-coding RNA (lncRNA) complexes. We next plan to build a large library of RNA fragments in order to extend this approach to partially structured RNA molecules, and in the longer term we aim to apply this approach to modeling flexible peptide-protein interactions in a similar way.

CARMEN Project-Team

4. Application Domains

4.1. Scientific context: the LIRYC

The University Hospital of Bordeaux (*CHU de Bordeaux*) is equipped with a specialized cardiology hospital, the *Hôpital Cardiologique du Haut-Lévêque*, where the group of Professor Michel Haïssaguerre has established itself as a global leader in the field of cardiac electrophysiology [53], [52], [45]. Their discoveries in the area of atrial fibrillation and sudden cardiac death syndromes are widely acclaimed, and the group is a national and international referral center for treatment of cardiac arrhythmia. Thus the group also sees large numbers of patients with rare cardiac diseases.

In 2011 the group has won the competition for a 40 million euro *Investissements d'Avenir* grant for the establishment of IHU Liryc, an institute that combines clinical, experimental, and numerical research in the area of cardiac arrhythmia (http://ihu-liryc.fr). The institute works in all areas of modern cardiac electrophysiology: atrial arrhythmias, sudden death due to ventricular fibrillation, heart failure related to ventricular dyssynchrony, and metabolic disorders. It is recognized as one of the most important centers worldwide in this area.

The Carmen team was founded to partner with IHU Liryc. We bring applied mathematics and scientific computing closer to experimental and clinical cardiac electrophysiology. In collaboration with experimental and clinical researchers at Liry we work to enhance fundamental knowledge of the normal and abnormal cardiac electrical activity and of the patterns of the electrocardiogram, and we develop new simulation tools for training, biological, and clinical applications.

4.2. Basic experimental electrophysiology

Our modeling is carried out in coordination with the experimental teams from IHU Liryc. It help to write new concepts concerning the multiscale organisation of the cardiac action potentials that will serve our understanding in many electrical pathologies. For example, we model the structural heterogeneities at the cellular scale [28], and at an intermediate scale between the cellular and tissue scales.

At the atrial level, we apply our models to understand the mechanisms of complex arrythmias and the relation with the heterogeneities at the insertion of the pulmonary veins. We will model the heterogeneities specific to the atria, like fibrosis or fatty infiltration [51]. These heterogeneities ar thought to play a major role in the development of atrial fibrillation.

At the ventricular level, we focus on (1) modeling the complex coupling between the Purkinje network and the ventricles and (2) modeling the heteogeneities related to the complex organization and disorganization of the myocytes and fibroblasts. Point (1) is supposed to play a major role in sudden cardiac death and point (2) is important in the study of infarct scars for instance.

4.3. Clinical electrophysiology

Treatment of cardiac arrhythmia is possible by pharmacological means, by implantation of pacemakers and defibrillators, and by curative ablation of diseased tissue by local heating or freezing. In particular the ablative therapies create challenges that can be addressed by numerical means. Cardiologists would like to know, preferably by noninvasive means, where an arrhythima originates and by what mechanism it is sustained.

We address this issue in the first place using inverse models, which attempt to estimate the cardiac activity from a (high-density) electrocardiogram. A new project aims at performing this estimation on-site in the catheterization laboratory and presenting the results, together with the cardiac anatomy, on the screen that the cardiologist uses to monitor the catheter positions [25].

An important prerequisite for this kind of interventions and for inverse modeling is the creation of anatomical models from imaging data. The Carmen team contributes to better and more efficient segmentation and meshing through the IDAM project.

CASTOR Project-Team

4. Application Domains

4.1. Magnetic confinment fusion

The main application domain is magnetic confined fusion. A part of the work is actually used on european Tokamaks like JET, WEST and ITER.

COFFEE Project-Team

4. Application Domains

4.1. Porous Media

Our research focuses on the numerical modeling of multiphase porous media flows accounting for complex geology and for nonlinear and multi-physics couplings. It is applied to various problems in the field of energy such as the simulation of geothermal systems in collaboration with BRGM, of nuclear waste repositories in collaboration with Andra, and of oil and gas recovery in collaboration with Total. Our research directions include the development of advanced numerical schemes adapted to polyhedral meshes and highly heterogeneous media in order to represent more accurately complex geologies. A special focus is made on the modeling of multiphase flows in network of faults or fractures represented as interfaces of co-dimension one coupled to the surrounding matrix. We also investigate nonlinear solvers adapted to the nonlinear couplings between gravity, capillary and viscous forces in highly heterogeneous porous media. In the same line, we study new domain decomposition algorithms to couple non-isothermal compositional liquid gas flows in a porous medium with free gas flows occurring at the interface between the ventilation gallery and the nuclear waste repository or between a geothermal reservoir and the atmosphere.

4.2. Particulate and mixture flows

We investigate fluid mechanics models referred to as "multi–fluids" flows. A large part of our activity is more specifically concerned with the case where a disperse phase interacts with a dense phase. Such flows arise in numerous applications, like for pollutant transport and dispersion, the combustion of fuel particles in air, the modelling of fluidized beds, the dynamic of sprays and in particular biosprays with medical applications, engine fine particles emission... There are many possible modelings of such flows: microscopic models where the two phases occupy distinct domains and where the coupling arises through intricate interface conditions; macroscopic models which are of hydrodynamic (multiphase) type, involving non standard state laws, possibly with non conservative terms, and the so–called mesoscopic models. The latter are based on Eulerian–Lagrangian description where the disperse phase is described by a particle distribution function in phase space. Following this path we are led to a Vlasov-like equation coupled to a system describing the evolution of the dense phase that is either the Euler or the Navier-Stokes equations. It turns out that the leading effect in such models is the drag force. However, the role of other terms, of more or less phenomenological nature, deserves to be discussed (close packing terms, lift term, Basset force...). Of course the fluid/kinetic model is interesting in itself and needs further analysis and dedicated numerical schemes. In particular, in collaboration with the Atomic Energy Commission (CEA), we have proposed a semi-Lagrangian scheme for the simulation of particulate flows, extending the framework established in plasma physics to such flows.

We also think it is worthwhile to identify hydrodynamic regimes: it leads to discuss hierarchies of coupled hydrodynamic systems, the nature of which could be quite intriguing and original, while they share some common features of the porous media problems. We are particularly interested in revisiting the modeling of mixture flows through the viewpoint of kinetic models and hydrodynamic regimes. We propose to revisit the derivation of new mixture models, generalizing Kazhikov-Smagulov equations, through hydrodynamic asymptotics. The model is of "hybrid" type in the sense that the constraint reduces to the standard incompressibility condition when the disperse phase is absent, while it involves derivatives of the particle volume fraction when the disperse phase is present.

4.3. Biological degradation, biofilms formation and algae proliferation

Members of the team have started an original research program devoted to biofilms formation and algae proliferation. We started working on this subject through a collaboration with Roberto Natalini and a group of experts in Firenze interested in preventing damages on historical monuments. It is also motivated by Ostreopsis proliferation in the Mediterranean Sea. The multidisciplinary character of this research relies on discussions with researchers of the Oceanography Laboratory in Villefranche-sur-Mer, a leading marine research unit, and the Inria team BIOCORE, led by J-L Gouzé. This research was supported by a ANR-project, led by M. Ribot, and it was the main topic of the PhD thesis of B. Polizzi.

DRACULA Project-Team (section vide)

<p align="center" style="color:red">DYLISS Project-Team</p>

4. Application Domains

4.1. Application fields in biology

Our methods are applied in several fields of molecular biology.

Our main application field is **marine biology**, as it is a transversal field with respect to issues in integrative biology, dynamical systems and sequence analysis. Our main collaborators work at the Station Biologique de Roscoff. We are strongly involved in the study of brown algae: the *meneco, memap and memerge* tools were designed to realize a complete reconstruction of metabolic networks for non-benchmark species [97], [69]. On the same application model, the pattern discovery tool *protomata learner* combined with supervised bi-clustering based on formal concept analysis allows for the classification of sub-families of specific proteins [67]. The same tool also allowed us to gain a better understanding of cyanobacteria proteins [4]. At the larger level of 4D structures, classification technics have also allowed us to introduce new methods for the characterization of viruses in marine metagenomic sample [72]. Finally, in dynamical systems, we use asymptotic analysis (tool *pogg*) to decipher the initiation of sea urchin translation [51] [95]. We are currently involved in two new applications in this domain: the team participates to a Inria Project Lab program with the Biocore and Ange Inria teams, focused on the understanding on green micro-algae; and we are involved in the deciphering of phytoplancton variability at the system biology level in collaboration with the Station Biologique de Roscoff (ANR Samosa).

In **micro-biology**, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile (funded by CMM, CRG and Inria-Chile). In order to elucidate the main characteristics of these bacteria, we develop efficient methods to identify the main groups of regulators for their specific response in their living environment. To that purpose, we use constraints-based modeling and combinatorial optimization. The integrative biology tools *meneco bioquali, ingranalysis, shogen, lombarde* were designed in this context [77]. In 2016, two applications focused on the study of extremophile consortium of bacteria have been performed with these tools [61], [38]. In parallel, in collaboration with Ifremer (Brest), we have conducted similar work to decipher protein-protein interactions within archebacteria [96]. Our sequence analysis tool (*logol*) allowed us to build and maintain a very expressive CRISPR database [10] [50].

Similarly, in **environmental sciences**, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues. In collaboration with researchers from Inra/Pegase laboratory, we develop methods to distinguish the response of breeding animals to different diaries or treatments [37] and characterize upstream transcriptional regulators [59], applied to porks [74], [75], [76]. Semantic-based analysis was useful for interpreting differences of gene expression in pork meat [79].

In addition, constraints-based programming also allows us to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants [85], [120]. This was performed in collaboration with Inra/Igepp. This paved the way to the recent research track initiated in the team about integration of heterogeneous data with RDF-technologies (see AskOmics software) [84], [70] and about graph-compression (see powergrasp software).

In **bio-medical applications**, we focus our attention on the confrontation of large-scale measurements with large-scale knowledge repositories about regulation pathways such as Transpath, PID or pathway commons. In collaboration with Institut Curie, we have studied the Ewing Sarcoma regulation network to test the capability of our tool *bioquali* to accurately correct and predict a large-scale network behavior [47]. Our ongoing studies in this field focus on the exhaustive learning of discrete dynamical networks matching with experimental data, as a case study for modeling experimental design with constraints-based approaches. To that purpose, we collaborate with J. Saez Rodriguez group at EBI [115] and N. Theret group at Inserm/Irset (Rennes) [40]. The dynamical system tools *caspo and cadbiom* were designed within these collaborations. Ongoing studies

focus on the understanding of the metabolism of xenobiotics (mecagenotox program) and the filtering of sets of regulatory compounds within large-scale signaling network (TGFSysBio project).

ERABLE Project-Team

4. Application Domains

4.1. Biology & Health

The main areas of application of ERABLE are: (1) biology understood in its more general sense, with a special focus on symbiosis and on intracellular interactions, and (2) health with a special emphasis for now on infectious diseases, rare diseases, and cancer.

FLUMINANCE Project-Team

4. Application Domains

4.1. Introduction

By designing new approaches for the analysis of fluid-image sequences the FLUMINANCE group aims at contributing to several application domains of great interest for the community and in which the analysis of complex fluid flows plays a central role. The group focuses mainly on two broad application domains:

- Environmental sciences;
- Experimental fluid mechanics and industrial flows.

We detail hereafter these two application domains.

4.2. Environmental sciences

The first huge application domain concerns all the sciences that aim at observing the biosphere evolution such as meteorology, climatology or oceanography but also remote sensing study for the monitoring of meteorological events or human activities consequences. For all these domains image analysis is a practical and unique tool to *observe, detect, measure, characterize or analyze* the evolution of physical parameters over a large domain. The design of generic image processing techniques for all these domains might offer practical software tools to measure precisely the evolution of fluid flows for weather forecasting or climatology studies. It might also offer possibilities of close surveillance of human and natural activities in sensible areas such as forests, river edges, and valley in order to monitor pollution, floods or fire. The need in terms of local weather forecasting, risk prevention, or local climate change is becoming crucial for our tomorrow's life. At a more local scale, image sensors may also be of major utility to analyze precisely the effect of air curtains for safe packaging in agro-industrial.

4.3. Experimental fluid mechanics and industrial flows

In the domain of **experimental fluid mechanics**, the visualization of fluid flows plays a major role, especially for turbulence study since high frequency imaging has been made currently available. Together with analysis of turbulence at different scales, one of the major goals pursued at the moment by many scientists and engineers consists in studying the ability to manipulate a flow to induce a desired change. This is of huge technological importance to enhance or inhibit mixing in shear flows, improve energetic efficiency or control the physical effects of strain and stresses. This is for instance of particular interest for:

- military applications, for example to limit the infra-red signatures of fighter aircraft;
- aeronautics and transportation, to limit fuel consumption by controlling drag and lift effects of turbulence and boundary layer behavior;
- industrial applications, for example to monitor flowing, melting, mixing or swelling of processed materials, or preserve manufactured products from contamination by airborne pollutants, or in industrial chemistry to increase chemical reactions by acting on turbulence phenomena.

.

GALEN Project-Team

4. Application Domains

4.1. Breast tomosynthesis

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet, Maissa Sghaier (collaboration G. Palma, GE Healthcare)

Breast cancer is the most frequently diagnosed cancer for women. Mammography is the most used imagery tool for detecting and diagnosing this type of cancer. Since it consists of a 2D projection method, this technique is sensitive to geometrical limitations such as the superimposition of tissues which may reduce the visibility of lesions or make even appear false structures which are interpreted by radiologists as suspicious signs. Digital breast tomosynthesis allows these limitations to be circumvented. This technique is grounded on the acquisition of a set of projections with a limited angle view. Then, a 3D estimation of the sensed object is performed from this set of projections, so reducing the overlap of structures and improving the visibility and detectability of lesions possibly present in the breast. The objective of our work is to develop a high quality reconstruction methodology where the full pipeline of data processing will be modeled.

4.2. Inference of gene regulatory networks

Participants: Jean-Christophe Pesquet (collaboration A. Pirayre and L. Duval, IFPEN)

The discovery of novel gene regulatory processes improves the understanding of cell phenotypic responses to external stimuli for many biological applications, such as medicine, environment or biotechnologies. To this purpose, transcriptomic data are generated and analyzed from DNA microarrays or more recently RNAseq experiments. They consist in genetic expression level sequences obtained for all genes of a studied organism placed in dierent living conditions. From these data, gene regulation mechanisms can be recovered by revealing topological links encoded in graphs. In regulatory graphs, nodes correspond to genes. A link between two nodes is identified if a regulation relationship exists between the two corresponding genes. In our work, we propose to address this network inference problem with recently developed techniques pertaining to graph optimization. Given all the pairwise gene regulation information available, we propose to determine the presence of edges in the considered GRN by adopting an energy optimization formulation integrating additional constraints. Either biological (information about gene interactions) or structural (information about node connectivity) a priori are considered to restrict the space of possible solutions. Different priors lead to different properties of the global cost function, for which various optimization strategies, either discrete and continuous, can be applied.

4.3. Lung Tumor Detection and Characterization

Participants: Evgenios Kornaropoulos, Evangelia Zacharaki, Nikos Paragios

The use of Diffusion Weighted MR Imaging (DWI) is investigated as an alternative tool to radiologists for tumor detection, tumor characterization, distinguishing tumor tissue from non-tumor tissue, and monitoring and predicting treatment response. In collaboration with Hôpitaux Universitaires Henri-Mondor in Paris, France and Chang Gung Memorial Hospital – Linkou in Taipei, Taiwan we investigate the use of modelbased methods of 3D image registration, clustering and segmentation towards the development of a framework for automatic interpretation of images, and in particular extraction of meaningful biomarkers in aggressive lymphomas.

4.4. Protein function prediction

Participants: Evangelia Zacharaki, Nikos Paragios (in collaboration with D. Vlachakis, University of Patras, Greece)

The massive expansion of the worldwide Protein Data Bank (PDB) provides new opportunities for computational approaches which can learn from available data and extrapolate the knowledge into new coming instances. The aim of our work was to exploit experimentally acquired structural information of enzymes through machine learning techniques in order to produce models that predict enzymatic function.

4.5. Imaging biomarkers for chronic lung diseases

Participants: Guillaume Chassagnon, Evangelia Zacharaki, Maria Vakalopoulou, Nikos Paragios
Diagnosis and staging of chronic lung diseases is a major challenge for both patient care and approval of new treatments. Among imaging techniques, computed tomography (CT) is the gold standard for in vivo morphological assessment of lung parenchyma currently offering the highest spatial resolution in chronic lung diseases. Although CT is widely used its optimal use in clinical practice and as an endpoint in clinical trials remains controversial. Our goal is to develop quantitative imaging biomarkers that allow (i) severity assessment (based on the correlation to functional and clinical data) and (ii) monitoring the disease progression. In the current analysis we focus on scleroderma and cystic fibrosis as models for restrictive and obstructive lung disease, respectively. Two different approaches are investigated: disease assessment by histogram or texture analysis and assessment of the regional lung elasticity through deformable registration. This work is in collaboration with the Department of Radiology, Cochin Hospital, Paris.

4.6. Co-segmentation and Co-registration of Subcortical Brain Structures

Participants: Enzo Ferrante, Nikos Paragios, Iasonas Kokkinos
New algorithms to perform co-segmentation and co-registration of subcortical brain structures on MRI images were investigated in collaboration with Ecole Polytechnique de Montreal and the Sainte-Justine Hospital Research Center from Montreal. Brain subcortical structures are involved in different neurodegenerative and neuropsychiatric disorders, including schizophrenia, Alzheimers disease, attention deficit, and subtypes of epilepsy. Segmenting these parts of the brain enables a physician to extract indicators, facilitating their quantitative analysis and characterization. We are investigating how estimated maps of semantic labels (obtained using machine learning techniques) can be used as a surrogate for unlabelled data. We are exploring how to combine them with multi-population deformable registration to improve both alignment and segmentation of these challenging brain structures.

4.7. Restoration of old video archives

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet (collaboration F. Abboud, WITBE, J.-H. Chenot and L. Laborelli, INA)
The last century has witnessed an explosion in the amount of video data stored with holders such as the National Audiovisual Institute whose mission is to preserve and promote the content of French broadcast programs. the cultural impact of these records, their value is increased due to commercial reexploitation through recent visual media. However, the perceived quality of the old data fails to satisfy the current public demand. The purpose of our work is to propose new methods for restoring video sequences supplied from television archive documents, using modern optimization techniques with proven convergence properties.

GENSCALE Project-Team

4. Application Domains

4.1. Introduction

Today, sequencing data are intensively used in many life science projects. The methodologies developed by the GenScale group are generic approaches that can be applied to a large panel of domains such as health, agronomy or environment areas. The next sections briefly describe examples of our activity in these different domains.

4.2. Health

Genetic and cancer disease diagnostic: Genetic diseases are caused by some particular mutations in the genomes that alter important cell processes. Similarly, cancer comes from changes in the DNA molecules that alter cell behavior, causing uncontrollable growth and malignancy. Pointing out genes with mutations helps in identifying the disease and in prescribing the right drug. Thus, DNA from individual patients is sequenced and the aim is to detect potential mutations that may be linked to the patient disease. Today the bioinformatics analysis is mainly based on the detection of SNPs (Single Nucleotide Polymorphism) from a set of predefined target genes. Tomorrow, due to the decreasing cost of the sequencing process, bioinformatics analysis will scan the complete genome and report all kinds of mutations, including complex mutations such as large insertions or deletions, that could be associated with cancers.

Neurodegenerative disorders: The biological processes that lead from abnormal protein accumulation to neuronal loss and cognitive dysfunction is not fully understood. In this context, neuroimaging biomarkers and statistical methods to study large datasets play a pivotal role to better understand the pathophysiology of neurodegenerative disorders. The discovery of new anatomical biomarkers could thus have a major impact on clinical trials by allowing inclusion of patients at a very early stage, at which treatments are the most likely to be effective. Correlations with genetic variables can determine subgroups of patients with common anatomical and genetic characteristics.

4.3. Agronomy and Environment

Improving plant breeding: such projects aim at 1) identifying favorable alleles at loci contributing to phenotypic variation, 2) characterizing N-traits at the functional level and 3) providing robust multi-locus SNP-based predictors of the breeding value of agronomical traits under polygenic control. Underlying bioinformatics processing is the detection of informative zones (QTL) on the plant genomes.

Insect genomics: Insects represent major crop pests, justifying the need for control strategies to limit population outbreaks and the dissemination of plant viruses they frequently transmit. Several issues are investigated through the analysis and comparison of their genomes: understanding their phenotypic plasticity such as their reproduction mode changes, identifying the genomic sources of adaptation to their host plant and of ecological speciation, and understanding the relationships with their bacterial symbiotic communities.

Ocean biodiversity: The metagenomic analysis of seawater samples provides an original way to study the ecosystems of the oceans. Through the biodiversity analysis of different ocean spots, many biological questions can be addressed, such as the plankton biodiversity and their role, for example, in the CO_2 sequestration.

IBIS Project-Team (section vide)

<p style="text-align:center">**LEMON Team**</p>

4. Application Domains

4.1. Inland flow processes

4.1.1. *Shallow water models with porosity*

Simulating urban floods and free surface flows in wetlands requires considerable computational power. Two-dimensional shallow water models are needed. Capturing the relevant hydraulic detail often requires computational cell sizes smaller than one meter. For instance, meshing a complete urban area with a sufficient accuracy would require 10^6 to 10^8 cells, and simulating one second often requires several CPU seconds. This makes the use of such model for crisis management impossible. Similar issues arise when modelling wetlands and coastal lagoons, where large areas are often connected by an overwhelming number of narrow channels, obstructed by vegetation and a strongly variable bathymetry. Describing such channels with the level of detail required in a 2D model is impracticable. A new generation of models overcoming this issue has emerged over the last 20 years: porosity-based shallow water models. They are obtained by averaging the two-dimensional shallow water equations over large areas containing both water and a solid phase [35]. The size of a computational cell can be increased by a factor 10 to 50 compared to a 2D shallow water model, with CPU times reduced by 2 to 3 orders of magnitude [53]. While the research on porosity-based shallow water models has accelerated over the past decade [48], [64], [68], [46], [45], [53], [78], [79], [74], [47], a number of research issues remain pending.

4.1.2. *Forcing*

Reproducing optimally realistic spatio-temporal rainfall fields is of salient importance to the forcing of hydro-dynamic models. This challenging task requires combining intense, usual and dry weather events. Far from being straightforward, this combination of extreme and non-extreme scenarios requires a realistic modelling of the transitions between normal and extreme periods. [57] have proposed in a univariate framework a statistical model that can serve as a generator and that takes into account low, moderate and intense precipitation. In the same vein, [75] developed a bivariate model. However, its extension to a spatial framework remains a challenge. Existing spatial precipitation stochastic generators are generally based on Gaussian spatial processes [22], [55], that are not adapted to generate extreme rainfall events. Recent advances in spatio-temporal extremes modelling based on generalized Pareto processes [39], [71] and semi-parametric simulation techniques [28] are very promising and could form the base for relevant developments in our framework.

4.1.3. *Inland hydrobiological systems*

Water bodies such as lakes or coastal lagoons (possibly connected to the sea) located in high human activity areas are subject to various kinds of stress such as industrial pollution, high water demand or bacterial blooms caused by freshwater over-enrichment. For obvious environmental reasons, these water resources have to be protected, hence the need to better understand and possibly control such fragile ecosystems to eventually develop decision-making tools. From a modelling point of view, they share a common feature in that they all involve interacting biological and hydrological processes. According to [40], models may be classified into two main types: "minimal dynamic models" and "complex dynamic models". These two model types do not have the same objectives. While the former are more heuristic and rather depict the likelihood of considered processes, the latter are usually derived from fundamental laws of biochemistry or fluid dynamics. Of course, the latter necessitate much more computational resources than the former. In addition, controlling such complex systems (usually governed by PDEs) is by far more difficult that controlling the simpler ODE-driven command systems.
LEMON has already contributed both to the reduction of PDE models for the simulation of water confinement in coastal lagoons [41], [23] and to the improvement of ODE models in order to account for space-heterogeneity of bioremediation processes in water resources [21].

4.1.4. Parametrization

Numerical modelling requires data acquisition, both for model validation and for parameter assessment. Model benchmarking against laboratory experiments is an essential step and is essential to team's strategy. However, scale model experiments may have several drawbacks: *i)* experiments are very expensive and extremely time-consuming, *ii)* experiments cannot always be replicated, and measurement have precision and reliability limitations, *iii)* dimensional similarity (in terms of geometry and flow characteristic variables such as Froude or Reynolds numbers) cannot always be preserved.

An ideal way to obtain data would be to carry out in situ measurements. But this would be too costly at the scale of studied systems, not to mention the fact that field may become impracticable during flood periods.

Remote sensing data are becoming widely available with high spatial and temporal resolutions. Several recent studies have shown that flood extends can be extracted from optical or radar images [42], for example: to characterize the flood dynamics of great rivers [58], to monitor temporary ponds [69], but also to calibrate hydrodynamics models and assess roughness parameters [66], [49], [77].

Upscaled models developed in LEMON embed new parameters that reflect the statistical properties of the medium geometry. Two types of information are needed: the directional properties of the medium and its flow connectivity properties. New methods are thus to be developed to characterize such statistical properties from geographical data.

4.2. Marine and coastal systems

4.2.1. Multi-scale ocean modelling

In physical oceanography, all operational models - regardless of the scale they apply to - are derived from the complete equations of geophysical fluid dynamics. Depending on the considered process properties (nonlinearity, scale) and the available computational power, the original equations are adapted with some simplifying hypotheses. The reader can refer to [63], [56] for a hierarchical presentation of such models.

In the nearshore area, the hydrostatic approximation that is used is most large scales models (high sea) cannot be used without a massive loss of accuracy. In particular, shallow water models are inappropriate to describe the physical processes that occur in this zone (see Figure 1 above). This is why Boussinesq-type models are prefered: see [54]. They embed dispersive terms that allow for shoaling and other bathymetry effects. Since the pioneering works of Green and Naghdi (see [43]), numerous theoretical and numerical studies have been delivered by the "mathematical oceanography" community, more specifically in France (see the works of Lannes, Marche, Sainte-Marie, Bresch, etc.). The corresponding numerical models (BOSZ, WaveBox) must thus be integrated in any reasonable nearshore modelling platform.

However, these models cannot simply replace all previous models everywhere in the ocean: dispersive models are useless away from the shore and it is known that wave breaking cannot be simulated using Boussinesq-type equations. Hence the need to couple these models with others. Some work has been done in this direction with a multi-level nesting using software packages such as ROMS, but to the best of our knowledge, all the "boxes" rely on the same governing equations with different grid resolutions. A real coupling between different models is a more difficult task since different models may have different mathematical properties, as shown in the work by Blayo and Rousseau on shallow water modelling (see [1] and [24]).

4.2.2. Data-model interactions

An alternative to direct observations is the chaining of numerical models, which for instance represent the physic from offshore to coastal areas. Typically, output data from atmospheric and ocean circulation models are used as forcings for a wave model, which in turn feeds a littoral model. In the case of extreme events, their numerical simulation from physical models is generally unreachable. This is due to a lack of knowledge on boundary conditions and on their physical reliability for such extreme quantities. Based on numerical simulated data, an alternative is to use statistical approaches. [28] proposed such an approach. They first produced and studied a 52-year hindcast using the WW3 wave model [26], [29], [27], [72]. Then stemming from parts of the original work of [25], [44], [39], [28] proposed a semi-parametric approach which aims

to simulate extreme space-time waves processes to, in turn, force a littoral hazard model. Nevertheless their approach allows only a very small number of scenarios to be simulated.

<h1 style="text-align:center;color:red">LIFEWARE Project-Team</h1>

4. Application Domains

4.1. Preamble

Our collaborative work on biological applications is expected to serve as a basis for groundbreaking advances in cell functioning understanding, cell monitoring and control, and novel therapy design and optimization. Our collaborations with biologists are focused on **concrete biological questions**, and on the building of predictive models of biological systems to answer them. However, one important application of our research is the development of a **modeling software** for computational systems biology.

4.2. Modeling software for systems biology

Since 2002, we develop an open-source software environment for modeling and analyzing biochemical reaction systems. This software, called the Biochemical Abstract Machine (BIOCHAM), is compatible with SBML for importing and exporting models from repositories such as BioModels. It can perform a variety of static analyses, specify behaviors in Boolean or quantitative temporal logics, search parameter values satisfying temporal constraints, and make various simulations. While the primary reason of this development effort is to be able to **implement our ideas and experiment them quickly on a large scale**, BIOCHAM is used by other groups either for building models, for comparing techniques, or for teaching (see statistics in software section). BIOCHAM-WEB is a web application which makes it possible to use BIOCHAM without any installation. We plan to continue developing BIOCHAM for these different purposes and improve the software quality.

4.3. Couplings between the cell cycle and the circadian clock

Recent advances in cancer chronotherapy techniques support the evidence that there exist important links between the cell cycle and the circadian clock genes. One purpose for modeling these links is to better understand how to efficiently target malignant cells depending on the phase of the day and patient characterictics. These questions are at the heart of our collaboration with Franck Delaunay (CNRS Nice) and Francis Lévi (Univ. Warwick, GB, formerly INSERM Hopital Paul Brousse, Villejuif) and of our participation in the ANR HYCLOCK project and in the submitted EU H2020 C2SyM proposal, following the former EU EraNet Sysbio C5SYS and FP6 TEMPO projects. In the past, we developed a coupled model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints [0]. We now focus on the bidirectional coupling between the cell cycle and the circadian clock and expect to gain fundamental insights on this complex coupling from computational modeling and single-cell experiments.

4.4. Biosensor design and implementation in non-living protocells

In collaboration with Franck Molina (CNRS, Sys2Diag, Montpellier) and Jie-Hong Jiang (NTU, Taiwan) we ambition to apply our techniques to the design and implementation of biosensors in non-living vesicles for medical applications. Our approach is based on purely protein computation and on our ability to compile controllers and programs in biochemical reactions. The realization will be prototyped using a microfluidic device at CNRS Sys2Diag which will allow us to precisely control the size of the vesicles and the concentrations of the injected proteins. It is worth noting that the choice of non-living chassis, in contrast to living cells in synthetic biology, is particularly appealing for security considerations and compliance to forthcoming EU regulation.

---

[0]Elisabetta De Maria, François Fages, Aurélien Rizk, Sylvain Soliman. Design, Optimization, and Predictions of a Coupled Model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints. Theoretical Computer Science, 412(21):2108 2127, 2011.

M3DISIM Project-Team

4. Application Domains

4.1. Clinical applications

After several validation steps – based on clinical and experimental data – we have reached the point of having validated the heart model in a pre-clinical context where we have combined direct and inverse modeling in order to bring predictive answers on specific patient states. For example, we have demonstrated the predictive ability of our model to set up pacemaker devices for a specific patient in cardiac resynchronization therapies, see [11]. We have also used our parametric estimation procedure to provide a quantitative characterization of an infarct in a clinical experiment performed with pigs, see [1].

<div align="center" style="color:red">**MAGIQUE-3D Project-Team**</div>

4. Application Domains

4.1. Seismic Imaging

The main objective of modern seismic processing is to find the best representation of the subsurface that can fit the data recorded during the seismic acquisition survey. In this context, the seismic wave equation is the most appropriate mathematical model. Numerous research programs and related publications have been devoted to this equation. An acoustic representation is suitable if the waves propagate in a fluid. But the subsurface does not contain fluids only and the acoustic representation is not sufficient in the general case. Indeed the acoustic wave equation does not take some waves into account, for instance shear waves, turning waves or the multiples that are generated after several reflections at the interfaces between the different layers of the geological model. It is then necessary to consider a mathematical model that is more complex and resolution techniques that can model such waves. The elastic or viscoelastic wave equations are then reference models, but they are much more difficult to solve, in particular in the 3D case. Hence, we need to develop new high-performance approximation methods.

Reflection seismics is an indirect measurement technique that consists in recording echoes produced by the propagation of a seismic wave in a geological model. This wave is created artificially during seismic acquisition surveys. These echoes (i.e., reflections) are generated by the heterogeneities of the model. For instance, if the seismic wave propagates from a clay layer to sand, one will observe a sharp reflected signal in the seismic data recorded in the field. One then talks about reflection seismics if the wave is reflected at the interface between the two media, or talks about seismic refraction if the wave is transmitted along the interface. The arrival time of the echo enables one to locate the position of this transition, and the amplitude of the echo gives information on some physical parameters of the two geological media that are in contact. The first petroleum exploration surveys were performed at the beginning of the 1920's and for instance, the Orchard Salt Dome in Texas (USA) was discovered in 1924 by the seismic-reflection method.

4.2. Imaging complex media with ultrasonic waves

The acoustic behavior of heterogeneous or composite materials attracts considerable excitement. Indeed, their acoustic response may be extremely different from the single constituents responses. In particular, dispersions of resonators in a matrix are the object of large research efforts, both experimentally and theoretically. However it is still a challenge to dispose of numerical tools with sufficient abilities to deal with the simulation and imaging of such materials behavior. Indeed, not only acoustic simulations are very time-consuming, but they have to be performed on realistic enough solution domains, i.e. domains which capture well enough the structural features of the considered materials.

This collaboration with I2M, University of Bordeaux aims at addressing this type of challenges by developing numerical and experimental tools in order to understand the propagation of ultrasonic waves in complex media, image these media, and in the future, help design composite materials for industrial purposes.

4.3. Helioseismology

This collaboration with the Max Planck Institute for Solar System, Göttingen, Germany, which started in 2014, aims at designing efficient numerical methods for the wave propagation problems that arise in helioseismology in the context of inverse problems. The final goal is to retrieve information about the structure of the Sun i.e. inner properties such as density or pressure via the inversion of a wave propagation problem. Acoustic waves propagate inside the Sun which, in a first approximation and regarding the time scales of physical phenomena, can be considered as a moving fluid medium with constant velocity of motion. Some other simplifications lead to computational saving, such as supposing a radial or axisymmetric geometry of

the Sun. Aeroacoustic equations must be adapted and efficiently solved in this context, this has been done in the finite elements code Montjoie. In other situations, a full 3D simulation is required and demands large computational resources. Ultimately, we aim at modeling the coupling with gravity potential and electromagnetic waves (MHD equations) in order to be able to better understand Sun spots.

MAMBA Project-Team (section vide)

MATHNEURO Team (section vide)

MIMESIS Team

4. Application Domains

4.1. Surgical Training

Virtual training prevents medical students from early manipulation of real patients. The development of simulation used for medical training usually requires important computational power, since realistic behaviours are key to deliver a high-fidelity experience to the trainee. Further, the quality of interaction with the simulator (usually via visual and haptic rendering) is also of a paramount importance. All these constraints make the development of training systems time-consuming thus limiting the deployment of virtual simulators in standard medical curriculum.

4.2. Pre-operative Planning

Beyond training, clinicians ask for innovative tools that can assist them in the pre-operative planning of an intervention. Using the patient information acquired before the operation, physics-based simulations allow to simulate the effect of a therapy with no risk to the patient. The clinicians can thus virtually assess different strategies and select the optimal procedure. Compared to a training simulation, a planning system requires a high accuracy to ensure reliability. Constrained by the time elapsed between the preoperative acquisition and the intervention, the computation must also be efficient.

4.3. Intra-operative Navigation

Besides the surgery training and planning, another major need from clinicians is surgical guidance. While the practician is performing the operation, a guidance system provides enriched visual feedback. This is especially useful with the emergence of minimally invasive surgery (MIS) where the visual information is often strongly limited. It can be used for example to avoid critical area such as vessels or to highlight the position of a tumour during its resection. In the MIS technique, the clinician does not interact with organs directly as in the open surgery, but manipulates instruments inserted through trocars placed in small incisions in the wall of the abdominal cavity. The surgeon can observe these instruments on a display showing a video stream captured by an endoscopic camera inserted through the navel. The main advantage of the method resides in reducing pain and time recovery, in addition to reducing bleeding and risks of infection. However, from a surgical standpoint, the procedure is quite complex since the field of view is considerably reduced and the direct manipulation of organs is not possible.

<div align="center">

MNEMOSYNE Project-Team

</div>

4. Application Domains

4.1. Overview

One of the most original specificity of our team is that it is part of a laboratory in Neuroscience (with a large spectrum of activity from the molecule to the behavior), focused on neurodegenerative diseases and consequently working in tight collaboration with the medical domain. As a consequence, neuroscientists and the medical world are considered as the primary end-users of our researches. Beyond data and signal analysis where our expertise in machine learning may be possibly useful, our interactions are mainly centered on the exploitation of our models. They will be classically regarded as a way to validate biological assumptions and to generate new hypotheses to be investigated in the living. Our macroscopic models and their implementation in autonomous robots will allow an analysis at the behavioral level and will propose a systemic framework, the interpretation of which will meet aetiological analysis in the medical domain and interpretation of intelligent behavior in cognitive neuroscience.

The study of neurodegenerative diseases is targeted because they match the phenomena we model. Particularly, the Parkinson disease results from the death of dopaminergic cells in the basal ganglia, one of the main systems that we are modeling. The Alzheimer disease also results from the loss of neurons, in several cortical and extracortical regions. The variety of these regions, together with large mnesic and cognitive deficits, require a systemic view of the cerebral architecture and associated functions, very consistent with our approach.

Of course, numerical sciences are also impacted by our researches, at several levels. At a global level, we will propose new control architectures aimed at providing a higher degree of autonomy to robots, as well as machine learning algorithms working in more realistic environment. More specifically, our focus on some cognitive functions in closed loop with a real environment will address currently open problems. This is obviously the case for planning and decision making; this is particularly the case for the domain of affective computing, since motivational characteristics arising from the design of an artificial physiology allow to consider not only cold rational cognition but also hot emotional cognition. The association of both kinds of cognition is undoublty an innovative way to create more realistic intelligent systems but also to elaborate more natural interfaces between these systems and human users.

At last, we think that our activities in well-founded distributed computations and high performance computing are not just intended to help us design large scale systems. We also think that we are working here at the core of informatics and, accordingly, that we could transfer some fundamental results in this domain.

<p style="text-align:center;color:red;font-weight:bold;">MONC Project-Team</p>

4. Application Domains

4.1. Tumor growth monitoring and therapeutic evaluation

Each type of cancer is different and requires an adequate model. More specifically, we are currently working on the following diseases:

- Glioma (brain tumors) of various grades,
- Metastases to the lung, liver and brain from various organs,
- Soft-tissue sarcoma,
- Kidney cancer and its metastases,
- EGFR-mutated lung cancer.

In this context our application domains are

- Image-driven patient-specific simulations of tumor growth and treatments,
- Parameter estimation and data assimilation of medical images.

4.2. Biophysical therapies

- Modeling of electrochemotherapy on biological and clinical scales.
- Evaluation of radiotherapy and radiofrequency ablation.

4.3. In-vitro and animals experimentations in oncology

- Theoretical biology of the metastatic process: dynamics of a population of tumors in mutual interactions, dormancy, pre-metastatic and metastatic niche, quantification of metastatic potential and differential effects of anti-angiogenic therapies on primary tumor and metastases.
- Mathematical models for preclinical cancer research: description and prediction of tumor growth and metastatic development, effect of anti-cancerous therapies.

MORPHEME Project-Team (section vide)

<div align="center">**MYCENAE Project-Team**</div>

4. Application Domains

4.1. Introduction

MYCENAE addresses rather "upstream" questions in neuroendocrinology and neuroscience. Nevertheless, MYCENAE's expected results can contribute to more applied issues in these fields, mainly by helping understand the mechanisms underlying physiological and pathological processes and also by designing new concepts for biomedical data analysis. MYCENAE thematics are related to societal issues concerning endocrine disruptors, reproductive biotechnologies, and neurological diseases, especially in case of pathological synchronizations encountered in epilepsy and Parkinson's disease.

4.2. Neuroendocrinology and Neuroscience

We are interested in the complex dynamical processes arising within neuroendocrine axes, with a special focus on the reproductive (hypothalamo-pituitary-gonadal) axis. This axis can be considered as the paragon of neuroendocrine axes, since it both concentrates all remarkable dynamics that can be exhibited by these axes and owns its unique specificities, as gonads are the only organs that host germ cells. Since, in neuroendocrine axes, neural systems are embedded within endocrine feedback loops and interact with peripheral organs, one also needs to get interested in the peripheral dynamics to be able to "close the loop" and account for the effect of peripheral inputs on neural dynamics. In the case of the HPG axis, these dynamics are especially complex, because they involve developmental processes that occur even in adult organisms and combine the glandular function of the gonads with their gametogenic function.

Neuroendocrinology is thus a scientific field at the interface between Neuroscience, Endocrinology and Physiology (and even of Developmental Biology in the case of the HPG axis). On a neuroscience ground, mathematical neuroendocrinology is specifically interested in endocrine neurons, which have the uncommon ability of secreting neurohormones into the blood stream. Neuroendocrine networks are characterized by the emergence of very slow rhythms (on the order of an hour), finite size effects due to their relative small number of neurons (on the order of a few thousands for the Gonadotropin-Releasing-Hormone network) and neuroanatomical particularities, that impact the way they can synchronize and desynchronize. On a physiological ground, gonadal cell biology raises specific cell biology issues on more than one account. First, the gonads are the only organs sheltering the germ cell lines (corresponding to oogenesis in ovaries and spermatogenesis in testes). Hence, the two modes of cell division, mitosis and meiosis are encountered in these tissues. Second, there are intricate interactions between the gonadal somatic cells (granulosa cells in the ovaries, sertoli cells in the testes) and the germ cells. Third, the control of gonadal cell populations is exerted within endocrine feedback loops involving both the hypothalamus and pituitary, which results naturally in multiscale population dynamics coupled with hormonally-controlled cell kinetics.

MYCENAE's research topics in mathematical neuroscience deal with complex oscillations, synchronization and plasticity.

We study (i) the emergence of network-level behaviors from individual dynamics of excitable cells (mainly neurons, but not exclusively, as the pituitary cells belong to the family of excitable cells): complete synchronization or synchronization of specific events, effect of the recruitment rate in the synchronization process, dependence on the neuro-anatomical and functional coupling properties; (ii) the control of the different possible configurations of the network depending on external (e.g. daylength) and/or internal inputs (e.g. metabolic status), at the source of plasticity processes in cognitive (vision learning) or neuroendocrine systems (differential sensitivity to gonadal steroids and peptides across the different steps of the reproductive life); (iii) the encoding of neuro-hormonal signals as complex oscillations, on the electrical, ionic (calcium dynamics) and secretory levels; and (iv) the decoding of these signals by their target neuronal or non-neuronal cells.

More recently, we have been interested into developmental biology issues in neurosciences: neurogenesis and brain development. The anatomical and functional organization of the nervous system, and especially the brain, is highly structured and tightly regulated. The surface of the cortex, its thickness, but also the size and shape of the brain areas associated to the different sensory or motor areas are very reliable quantities across different individuals. In collaboration with different teams of biologists, we develop and investigate models of the development of the brain, at different time and spatial scale.

The biological relevance of our modeling and model-based signal analysis approaches is grounded on our network of collaborations with teams of experimentalist biologists. In particular, we have long standing collaborations with the UMR 6175 (INRA-CNRS-Université François Rabelais-Haras Nationaux) "Physiologie de la Reproduction et des Comportements" that covers most our research topics in reproductive neuroendocrinology. We have especially close links with the Bingo (Integrative Biology of the ovary) and Bios (Biology and Bioinformatics of Signaling Systems) teams, which were partners of the REGATE LSIA. We have been jointly investigating issues relative to terminal or basal follicular development [7], [8], analysis of neurosecretory patterns [16] and modeling of GPCR (G-Protein Coupled Receptors) signaling networks [10]. We also have special links with the Center for Interdisciplinary Research in Biology (CIRB, Collège de France), headed by Alain Prochiantz, that help us get a better understanding of how the brain connectivity develops and how it is functionally organized. An instance of a recent collaborative work is the study of the organization of spatial frequencies in the primary visual cortex [34].

<p style="color:red; text-align:center">NEUROSYS Project-Team</p>

4. Application Domains

4.1. General remarks

The research directions of the team are motivated by general anesthesia that has attracted our attention in the last years. The following paragraphs explain in detail the motivation of our work on the four major phenomena of general anesthesia: loss of consciousness, immobility, amnesia and analgesia.

During general anesthesia, the electroencephalogram (EEG) on the scalp changes characteristically: increasing the anesthetic drug concentration the amplitudes of oscillations in the α-band ($\sim 8 - 12$Hz) and in the $\delta-$band ($2 - 8$Hz) increase amplitudes in frontal electrodes at low drug concentrations whereas the spectral power decreases in the $\gamma-$band ($\sim 20 - 60$Hz). This characteristic change in the power is the basis of today's EEG-monitors that assist the anesthetist in the control of the anesthesia depths of patients during surgery. However, the conventional monitors exhibit a large variability between the detected anesthetic depth and the real depth of patients. Moreover, a certain number of patients re-gain consciousness during surgery (about $1 - 2$ out of 1000) and a large percentage of patients suffer from diverse after-effects, such as nausea or long-lasting cognitive impairments such as partial amnesia (from days to weeks). Since surgery under general anesthesia is part of a hospital's everyday practice, a large number of patients suffer from these events every day. One reason for the lacking control of such disadvantageous effects is the dramatic lack of knowledge on what is going on in the brain during general anesthesia and a weak EEG-online monitoring system during anesthesia. Consequently, to improve the situation of patients during and after surgery and to develop improved anesthetic procedures or even drugs, research is necessary to learn more about the neural processes in the brain and develop new monitoring machines.

4.2. Level of consciousness

The EEG originates from coherent neural activity of populations in the cortex. Hence to understand better the characteristic power changes in EEG during anesthesia, it is necessary to study neural population dynamics subject to the concentration of anesthetic drugs and their action on receptors on the single neuron level. We study mathematical models which will be constrained by the signal features extracted from experimental data, such as EEG (data provided by Jamie Sleigh, university of Auckland and Christoph Destrieux, university of Tours), Local Field Potentials (data provided by Flavio Fröhlich, university of North Carolina - Chapel Hill) and behavior. The combination of model and analysis of experimental data provides the optimal framework to reveal new knowledge on the neural origin of behavioral features, such as the loss of consciousness or the uncontrolled gain of consciousness during surgery. For instance, modeling studies show that the characteristic changes of spectral power (second-order statistics) are not sufficient to deduce all underlying neural mechanisms. Consequently, additional higher-order statistical measures may provide additional insight into underlying neural mechanisms and may provide a novel marker for the loss of consciousness.

Moreover, the constant supervision of anesthetized patients in intensive care is a demanding task for the personnel in hospital practice. It is almost not possible to take care of a patient constantly and hence the todays' medicine demands monitoring devices that control automatically the level of anesthetic drugs based on the patients' neural activity (e.g., EEG). Brain-Computer-Interfaces (BCI) have already demonstrated their potential for the detection of consciousness in non-responsive patients. We will apply the data analysis techniques known in BCI to extract new markers for the depth of anesthesia. More specifically, for deeper anesthesia, auditory-evoked and Event-Related Desynchronization/Event-Related Synchronization (ERD/ERS) BCI could be used to better identify the state of consciousness in patients under anesthesia. In this context, we have established a first contact to the university of Würzburg. Another research direction will link intracranial EEG and scalp EEG by characterising micro-awake episodes during sleep.

4.3. Immobility

A research direction will be to take benefit of the relationship between the motor activity and anesthesia. Indeed, even if no movement is visually perceptible, a study by electroencephalographic recordings of brain activity in motor areas, quantifying the characteristics of amplitude and phase synchronization observed in the alpha and beta frequency bands, may reveal an intention of movement. This feature is important because it demonstrates that the patient is aware. Thus, we will develop an experimental protocol in collaboration with an anesthesiologist of the regional hospital on stimulating the median nerve at forearm level to track the evolution of the shape of the beta rebound in the motor cortex for various doses of the anesthetic agent.

4.4. Amnesia

Patients sometimes develop post-traumatic disorders associated with the surgery they underwent because they either woke up during the surgery or because the amnesiant effect of the general anesthesia was only partial, declarative memory being maintained in some unexplained cases. It is still unknown how memory can be maintained under general anesthesia and it needs to be investigated to improve the recovery from anesthesia and to avoid as much as possible post-traumatic disorders. To learn more about memory under anesthesia, we will focus our theoretical studies on the oscillation regimes observed in the hippocampus, mainly in the theta and gamma ranges, which are correlated with memory formation and retrieval.

4.5. Analgesia

One of the most important aspect in general anesthesia is the loss of pain. During surgery, it is very difficult to find out whether the anesthetized patient feels pain and hence will develop cognitive impairment after surgery. Today, the anesthesiologist knows and detects physiological signs of pain, such as sweat, colour of skin or spontaneous involuntary movements. However, more objective criteria based on EEG may assist the pain detection and hence improve the patients' situation. To this end, we analyze large sets of patient EEG-data observed during surgery and aim to extract EEG signal features of pain.

NUMED Project-Team (section vide)

<div align="center">

PARIETAL Project-Team

</div>

4. Application Domains

4.1. Cognitive neuroscience

4.1.1. *Macroscopic Functional cartography with functional Magnetic Resonance Imaging (fMRI)*

The brain as a highly structured organ, with both functional specialization and a complex network organization. While most of the knowledge historically comes from lesion studies and animal electophysiological recordings, the development of non-invasive imaging modalities, such as fMRI, has made it possible to study routinely high-level cognition in humans since the early 90's. This has opened major questions on the interplay between mind and brain , such as: How is the function of cortical territories constrained by anatomy (connectivity) ? How to assess the specificity of brain regions ? How can one characterize reliably inter-subject differences ?

4.1.2. *Analysis of brain Connectivity*

Functional connectivity is defined as the interaction structure that underlies brain function. Since the beginning of fMRI, it has been observed that remote regions sustain high correlation in their spontaneous activity, i.e. in the absence of a driving task. This means that the signals observed during resting-state define a signature of the connectivity of brain regions. The main interest of resting-state fMRI is that it provides easy-to-acquire functional markers that have recently been proved to be very powerful for population studies.

4.1.3. *Modeling of brain processes (MEG)*

While fMRI has been very useful in defining the function of regions at the mm scale, Magneto-encephalography (MEG) provides the other piece of the puzzle, namely temporal dynamics of brain activity, at the ms scale. MEG is also non-invasive. It makes it possible to keep track of precise schedule of mental operations and their interactions. It also opens the way toward a study of the rhythmic activity of the brain. On the other hand, the localization of brain activity with MEG entails the solution of a hard inverse problem.

4.1.4. *Current challenges in human neuroimaging (acquisition+analysis)*

Human neuroimaging targets two major goals: *i)* the study of neural responses involved in sensory, motor or cognitive functions, in relation to models from cognitive psychology, i.e. the identification of neurophysiological and neuroanatomical correlates of cognition; *ii)* the identification of markers in brain structure and function of neurological or psychiatric diseases. Both goals have to deal with a tension between

- the search for higher spatial [0] resolution to increase **spatial specificity** of brain signals, and clarify the nature (function and structure) of brain regions. This motivates efforts for high-field imaging and more efficient acquisitions, such as compressed sensing schemes, as well as better source localization methods from M/EEG data.

- the importance of inferring brain features with **population-level** validity, hence, contaminated with high variability within observed cohorts, which blurs the information at the population level and ultimately limits the spatial resolution of these observations.

---

[0]and to some extent, temporal, but for the sake of simplicity we focus here on spatial aspects.

Importantly, the signal-to-noise ratio (SNR) of the data remains limited due to both resolution improvements [0] and between-subject variability. Altogether, these factors have led to realize that results of neuroimaging studies were **statistically weak**, i.e. plagued with low power and leading to unreliable inference [54], and particularly so due to the typically number of subjects included in brain imaging studies (20 to 30, this number tends to increase [55]): this is at the core of the *neuroimaging reproducibility crisis*. This crisis is deeply related to a second issue, namely that only few neuroimaging datasets are publicly available, making it impossible to re-assess a posteriori the information conveyed by the data. Fortunately, the situation improves, lead by projects such as NeuroVault or OpenfMRI. A framework for integrating such datasets is however still missing.

---

[0]The SNR of the acquired signal is proportional to the voxel size, hence an improvement by a factor of 2 in image resolution along each dimension is payed by a factor of 8 in terms of SNR.

<p style="text-align:center; color:red">PLEIADE Team</p>

4. Application Domains

4.1. Genome and transcriptome annotation, to model function

Sequencing genomes and transcriptomes provides a picture of how a biological system can function, or does function under a given physiological condition. Simultaneous sequencing of a group of related organisms is now a routine procedure in biological laboratories for studying a behavior of interest, and provides a marvelous opportunity for building a comprehensive knowledge base of the relations between genomes. Key elements in mining these relations are: classifying the genes in related organisms and the reactions in their metabolic networks, recognizing the patterns that describe shared features, and highlighting specific differences.

PLEIADE will develops applications in comparative genomics of related organisms, using new mathematical tools for representing compactly, at different scales of difference, comparisons between related genomes. New methods based on computational geometry refine these comparisons. Compact representations can be stored, exchanged, and combined. They will form the basis of new simultaneous genome annotation methods, linked directly to abductive inference methods for building functional models of the organisms and their communities.

Our ambition in biotechnology is to permit the design of synthetic or genetically selected organisms at an abstract level, and guide the modification or assembly of a new genome. Our effort is focused on two main applications: genetic engineering and synthetic biology of oil-producing organisms (biofuels in CAER, palm oils), and improving and selecting starter microorganisms used in winemaking (collaboration with the ISVV and the BioLaffort company).

4.2. Molecular based systematics and taxonomy

Defining and recognizing myriads of species in biosphere has taken phenomenal energy over the past centuries and remains a major goal of Natural History. It is an iconic paradigm in pattern recognition (clustering has coevolved with numerical taxonomy many decades ago). Developments in evolution and molecular biology, as well as in data analysis, have over the past decades enabled a profound revolution, where species can be delimited and recognized by data analysis of sequences. We aim at proposing new tools, in the framework of E-science, which make possible (i) better exploration of the diversity in a given clade, and (ii) assignment of a place in these patterns for new, unknown organisms, using information provided by sets of sequences. This will require investment in data analysis, machine learning, and pattern recognition to deal with the volumes of data and their complexity.

One example of this project is about the diversity of trees in Amazonian forest, in collaboration with botanists in French Guiana. Protists (unicellular Eukaryots) are by far more diverse than plants, and far less known. Molecular exploration of Eukaryotes diversity is nowadays a standard in biodiversity studies. Data are available, through metagenomics, as an avalanche and make molecular diversity enter the domain of Big Data. Hence, an effort will be invested, in collaboration with other Inria teams (GenScale, HiePACS) for porting to HPC algorithms of pattern recognition and machine learning, or distance geometry, for these tools to be available as well in metagenomics. This will be developed first on diatoms (unicellular algae) in collaboration with INRA team at Thonon and University of Uppsala), on pathogens of tomato and grapewine, within an existing network, and on bacterial communities, in collaboration with University of Pau. For the latter, the studies will extend to correlations between molecular diversity and sets of traits and functions in the ecosystem.

4.3. Community ecology and population genetics

Community assembly models how species can assemble or diassemble to build stable or metastable communities. It has grown out of inventories of countable organisms. Using *metagenomics* one can produce molecular based inventories at rates never reached before. Most communities can be understood as pathways of carbon exchange, mostly in the form of sugar, between species. Even a plant cannot exist without carbon exchange with its rhizosphere. Two main routes for carbon exchange have been recognized: predation and parasitism. In predation, interactions–even if sometimes dramatic–may be loose and infrequent, whereas parasitism requires what Claude Combes has called intimate and sustainable interactions [21]. About one decade ago, some works [25] have proposed a comprehensive framework to link the studies of biodiversity with community assembly. This is still incipient research, connecting community ecology and biogeography.

We aim at developping graph-based models of co-occurence between species from NGS inventories in metagenomics, i.e. recognition of patterns in community assembly, and as a further layer to study links, if any, between diversity at different scales and community assemblies, starting from current, but oversimplified theories, where species assemble from a regional pool either randomly, as in neutral models, or by environmental filtering, as in niche modeling. We propose to study community assembly as a multiscale process between nested pools, both in tree communities in Amazonia, and diatom communities in freshwaters. This will be a step towards community genomics, which adds an ecological flavour to metagenomics.

Convergence between the processes that shape genetic diversity and community diversity–drift, selection, mutation/speciation and migration–has been noted for decades and is now a paradigm, establishing a continuous scale between levels of diversity patterns, beyond classical approaches based on iconic levels like species and populations. We will aim at deciphering diversity pattern along these gradients, connecting population and community genetics. Therefore, some key points must be adressed on reliability of tools.

Next-generation sequencing technologies are now an essential tool in population and community genomics, either for making evolutionary inferences or for developing SNPs for population genotyping analyses. Two problems are highlighted in the literature related to the use of those technologies for population genomics: variable sequence coverage and higher sequencing error in comparison to the Sanger sequencing technology. Methods are developed to develop unbiased estimates of key parameters, especially integrating sequencing errors [24]. An additional problem can be created when sequences are mapped on a reference sequence, either the sequenced species or an heterologous one, since paralogous genes are then considered to be the same physical position, creating a false signal of diversity [22]. Several approaches were proposed to correct for paralogy, either by working directly on the sequences issued from mapped reads [22] or by filtering detected SNPs. Finally, an increasingly popular method (RADseq) is used to develop SNP markers, but it was shown that using RADseq data to estimate diversity directly biases estimates [15]. Workflows to implement statistical methods that correct for diversity biases estimates now need an implementation for biologists.

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.

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.* [50], [55], [54] 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). One of the goal is to use our ECG simulator to address the inverse problem of electrocardiology. In collaboration with the Macs/M3disym project-team, we are interested in the electromechanical coupling in the myocardium. We are also interested in various clinical and industrial issues related to cardiac electrophysiology, in particular the simulation of experimental measurement of the field potential of cardiac stem cells in multi-electrode arrays.

SERENA Project-Team

4. Application Domains

4.1. Multiphase flows and transport of contaminants in the subsurface

- subsurface depollution after chemical leakage
- nuclear waste disposal in deep underground repositories
- geological sequestration of CO_2
- production of oil and gas

4.2. Complex Stokes and Navier–Stokes flows

- industrial risks in energy production (fission)

4.3. Energy production, sustainable use of resources

- simulation of shock waves impinging on deformable or fragmentable structures
- use of nets of rods for sustainable construction

4.4. Computational quantum chemistry

- guaranteed bounds for ground-state energy (eigenvalues) and ground-state density matrix (eigenvectors) in first-principle molecular simulation
- application to Laplace, Gross–Pitaevskii, Kohn–Sham, and Schrödinger models

<p style="text-align:center">**SERPICO Project-Team**</p>

4. Application Domains

4.1. Modeling and analysis of membrane transport and molecule trafficking at the single cell scale
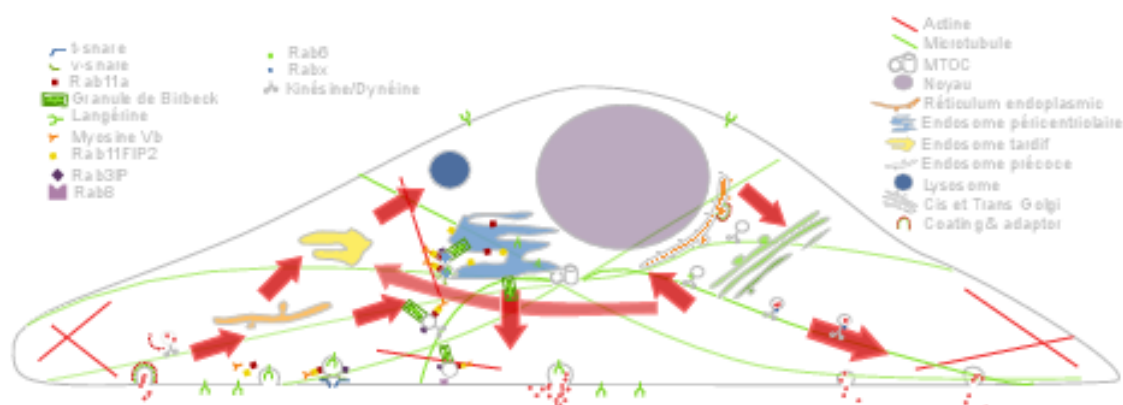
Figure 1. Cargo Langerin Trafficking controlled by Rab11A/Rab11FIP2/MyoVb platform.

In the past recent years, research carried at UMR 144 CNRS-Institut Curie ("Space Time imaging of Endomembranes and organelles Dynamics" team) contributed to a better understanding of the intracellular compartimentation of specialized model cells such as melanocytes and Langerhans cells, the components and structural events involved in the biogenesis of their specialized organelles: melanosomes and Birbeck granules, respectively. These studies have started to highlight: i/ multiple sorting and structural events involved in the biogenesis of these organelles; ii/ complexity of the endo-melanosomal network of these highly specialized cells; iii/ complex molecular architecture organizing and coordinating their dynamics; iv/ intracellular transport steps affected in genetic diseases, among which the Hermansky Pudlak syndrome (HPS) or involved in viral infection (HIV and Langerin in Langerhans cells).

In this context, the central aim of SERPICO is to understand how the different machineries of molecular components involved are interconnected and coordinated to generate such specialized structures. We need to address the following topics:

1. developing new bioimaging approaches to observe and statistically analyze such coordinated dynamics in live material;
2. correlating this statistically relevant spatiotemporal organization of protein networks with the biological architectures and at the ultrastructural level;
3. modeling intracellular transport of those reference biological complex systems and proposing new experimental plans in an iterative and virtuous circle;
4. managing and analyzing the workflow of image data obtained along different multidimensional microscopy modalities.

These studies are essential to unravel the complexity of the endomembrane system and how different machineries evolve together (e.g. see Fig. 1). They help to control cell organization and function at different scales through an integrative workflow of methodological and technological developments.

At long term, these studies will shed light on the cellular and molecular mechanisms underlying antigen presentation, viral infection or defense mechanisms, skin pigmentation, the pathogenesis of hereditary genetic disorders (lysosomal diseases, immune disorders) and on the mechanisms underlying cell transformation. Our methodological goal is also to link dynamics information obtained through diffraction limited light microscopy, eventually at a time regime compatible with live cell imaging. The overview of ultrastructural organization will be achieved by complementary electron microscopical methods. Image visualization and quantitative analysis are of course important and essential issues in this context.

4.2. Imaging and analysis of cytokskeleton dynamics during cell migration

The ability to migrate in space is among the most fundamental functions of eukaryotic cells and thus is one of the best-studied phenomena in biology. During embryonic development, cell movements result in a massive reorganization of the embryo, from a simple spherical ball of cells into a multi-layered organism; many of the cells at or near the surface of the embryo move to a new, more interior location. Moreover, inadequate or inappropriate migration of immune cells is also critically important for the delivery of protective immune responses to tissues and for wound healing. Finally, cell migration may facilitate the dissemination of tumor cells in blood and organs and eventually the formation of secondary tumors and metastases.

It has been established that the cytoskeleton, composed of actin filaments, microtubules and intermediate filaments (elongated structures with a diameter of a few dozens of nanometers), is essential for several cell mechanisms, including cell migration, cell division and molecule trafficking:

> i/ the actin filaments promote cell protrusion, adhesion and retraction;

> ii/ the microtubules are the support of molecule traffic and cell polarization;

> iii/ the intermediate filaments are hypothesized to control microtubule organization.

Nevertheless, the mechanical and chemical states of migrating cells under various external conditions remain largely unknown. In the last decade, high-resolution microscopy methods led to the discovery of novel aspects of cell migration. Most approaches and models are limited to migration in 2D, justified by the flatness of the cell-motile mechanisms. However, the mechanical patterns that govern migration in 2D models are often not essential for efficient migration in 3D. Accordingly, recent very challenging 3D models of cells moving on flat surfaces have begun to emerge. The key challenge, however, is to understand how a 3D motile cell crawls through the 3D extracellular matrix.

The objective of SERPICO is to develop high-end signal processing and computer vision tools to unfold the dynamical coordination of microtubules, actin filaments and intermediate filaments in 3D, involved in cell migration, cell division and molecule trafficking.

<p style="text-align:center; color:red;">**SISTM Project-Team**</p>

4. Application Domains

4.1. Systems Biology and Translational medicine

Biological and clinical researches have dramatically changed because of the technological advances, leading to the possibility of measuring much more biological quantities than previously. Clinical research studies can include now traditional measurements such as clinical status, but also thousands of cell populations, peptides, gene expressions for a given patient. This has facilitated the transfer of knowledge from basic to clinical science (from "bench side to bedside") and vice versa, a process often called "Translational medicine". However, the analysis of these large amounts of data needs specific methods, especially when one wants to have a global understanding of the information inherent to complex systems through an "integrative analysis". These systems like the immune system are complex because of many interactions within and between many levels (inside cells, between cells, in different tissues, in various species). This has led to a new field called "Systems biology" rapidly adapted to specific topics such as "Systems Immunology" [55], "Systems vaccinology" [52], "Systems medicine" [44]. From the statistician point of view, two main challenges appear: i) to deal with the massive amount of data ii) to find relevant models capturing observed behaviors.

4.2. The case of HIV immunology

The management of HIV infected patients and the control of the epidemics have been revolutionized by the availability of highly active antiretroviral therapies. Patients treated by these combinations of antiretrovirals have most often undetectable viral loads with an immune reconstitution leading to a survival which is nearly the same to uninfected individuals [48]. Hence, it has been demonstrated that early start of antiretroviral treatments may be good for individual patients as well as for the control of the HIV epidemics (by reducing the transmission from infected people) [43]. However, the implementation of such strategy is difficult especially in developing countries. Some HIV infected individuals do not tolerate antiretroviral regimen or did not reconstitute their immune system. Therefore, vaccine and other immune interventions are required. Many vaccine candidates as well as other immune interventions (IL7, IL15) are currently evaluated. The challenges here are multiple because the effects of these interventions on the immune system are not fully understood, there are no good surrogate markers although the number of measured markers has exponentially increased. Hence, HIV clinical epidemiology has also entered in the era of Big Data because of the very deep evaluation at individual level leading to a huge amount of complex data, repeated over time, even in clinical trials that includes a small number of subjects.

4.3. The case of Ebola vaccine development

In response to the recent outbreak of Ebola virus disease in West Africa, the clinical development of some candidate to Ebola vaccine has been accelerated. Several vectors, mostly encoding glycoprotein of the virus, were tested in Phase I-II studies in order to assess their safety and immunogenicity. One of the main question of interest there is the antibody response induced by vaccination, as some non-human primates studies have shown protection against the virus when antibody levels were high enough. Although bridging studies still have to be developed, antibodies are thus considered as a criterium of interest. The challenge is then to evaluate the durability of the antibody response, whether it be at an individual or population level, in order to evaluate the impact of a vaccine strategy in case of an epidemic. Moreover, we are interested in the factors associated to this antibody response, and even more the other immune markers (from both innate and adaptative immune response) able to predict antibody levels. As those relationship are non-linear, sophisticated statistical and mathematical methods are developed in order to address these questions. A systems medicine approach using multidimensional immunogenicity data from clinical trials and statistical models can help to understand vaccine mechanisms and improve the selection of optimised vaccine strategies for clinical trials.

<p style="text-align:center">**STEEP Project-Team**</p>

4. Application Domains

4.1. Introduction

In the context described in the previous sections, we can distinguish two connected and complementary strategies for analyzing environmental pressures: a sectorial approach and a spatial one. The first one is more directly connected to ecological accounting, the second one has more direct relations to urban economy and land cover modelling. Let us start by describing the former.

4.2. Ecological accounting for sectorial pressure assessment

One of the major issues in the assessment of the long-term sustainability of urban areas is related to the concept of "imported sustainability". Cities bring in from the outside most of their material and energy resources, and reject to the outside the waste produced by their activity. The modern era has seen a dramatic increase in both volume and variety of these material flows and consumption as well as in distance of origin and destination of these flows, usually accompanied by a spectacular increase in the associated environmental impacts. A realistic assessment of the sustainability of urban areas requires to quantify both local and distant environmental impacts; greenhouse gas emissions are only one aspect of this question. Such an assessment brings to light the most relevant direct and indirect lines of action on these issues. In this respect, it is useful to introduce the alternative concepts of consumer versus producer responsibility (or point of view).

The producer point of view is the most useful to pinpoint relevant direct lines of actions on environmental pressures due to production. In other respects, any territory imports and exports goods and services from and to the rest of the world. The consumer point of view provides information on the indirect pressures associated with these exchanges, as production responds to a final demand. Tracking the various supply chains through the analysis of the structure of the local economy and its relations and dependencies to the external world allows us to identify critically important contributions to environmental pressures; this also enables us to define fair environmental indicators in order not to attribute environmental pressures to producers only (whose responsibility is the easier to quantify of the two). In this approach, the producer responsibility follows directly from the measurement of its energy and material uses, while the consumer responsibility is established indirectly through an allocation of the impacts of production to the final consumers, but this second mode of allocation is to some extent virtual and partly subjective. Four methods stand out:

- Material Flow Analysis (MFA)
- Input-Output Analysis (IOA)
- Life-Cycle Analysis (LCA)
- Ecological Footprint (EF)

Each of these is based on a well-defined structuring element: mass conservation for MFA, measure of industrial inter-dependencies for IOA, identification of all the steps from cradle to grave for LCA, measure of biocapacity demand for EF. The different methods have preferred areas of application. For example, EF is more relevant for analyzing primary production such as agricultural staples, wood, etc. IOA is more focused on whole industrial sectors, while LCA is geared towards end-user products, taken as functional units; finally, primary materials (such as metals), waste and emissions are more easily characterized through MFA. Methodological choices are driven by the type of question one needs to address, data availability and collection method and the spatial scales under consideration. Indeed, data can be used in two different ways: bottom-up or top-down. The bottom-up data is more precise, but in general precludes comprehensiveness; on the contrary, the top-down data is by nature more comprehensive, but is not suited for a detailed, fine-scale analysis of the results.

STEEP is pursuing its research program on this theme with three major goals: 1) Creating a comprehensive database enabling pressure analyses; 2) Developing methodologies and models resolving scaling issues, and developing algorithms allowing us to rigorously and automatically obtain adequate assessments; 3) Providing a synthetic analysis of environmental pressures associated to the major material flows, at various geographic levels (employment catchment area, *département* and *région*, for France), with the explicit aim of incorporating this type of information in the public decision process on environmental issues, via specifically designed decision-help procedures.

4.3. Urban economy and land use/land cover changes: assessment of spatial distributions of the pressures

The preceding section was focused on territorial metabolism, in particular on the analysis of supply chains. Here territories are examined with a more prominent emphasis on their spatial dimension, with attention to: the spatial distribution of local pressures previously identified (from a land use point of view), and the modeling of future land use and activity location (from an economic point of view). These two questions correspond to very different modeling strategies: the first one is more statistical in nature, extrapolating future land use from past evolution combined with global territory scenarios; the other one has a more fundamental flavor and focuses on an understanding of the processes driving urbanization. For this, we focus more precisely on the question of household and businesses choices of localization, as well as on spatial fluxes within the territory (transportation of goods and persons). The critical point here is to understand and manage urban sprawl and its environmental effects (GHG emission, loss of arable land, ecosystem fragmentation, and so on).

4.3.1. Land Use/Land Cover Change models (LUCC)

LUCC models are mostly used in environmental sciences, e.g. to evaluate the impact of climate change on agriculture, but they can also be used to analyze urban sprawl. There is a variety of models, static or dynamic, grid- or agent- based, local or global, etc., and with varying degrees of sophistication concerning spatio-temporal analysis or decision structures incorporated in the model.

The models of interest here are statistical in nature but spatially explicit. Following decades of development, they are robust, versatile and mature. In principle, agent-models have a larger potential for representing decision processes, but in practice this advantage results in a loss of universality of the models. Among the most well-known and most mature models, one can mention the CLUE family of models, DINAMIC, or LCM (Land Change Modeler. These models are well described in the literature, and will only be briefly presented here.

These models analyze change in land use in a statistical way; they are structured around three different modules:

- The first module determines the probability of change of pixels of the territory (pixels are typically tens to hundreds of meters in size).
- The second module defines the global changes between the various land uses of interest per time step (usually, a few years), based on global scenarios of evolution of the territory under study. These first two modules are independent of one another.
- The last module distributes changes of land use in an explicit manner, pixel per pixel, at each time step, on the basis of the information provided by the first two modules.

Probabilities of change are calibrated on past evolution, from the differences between two past maps of land use in the more favorable cases, or from a single map otherwise (under the assumption that the logic of occupation changes is the same as the logic of land use at this single date). Such changes are then characterized in a statistical way with the help of modeling variables identified by the modeler as having potential explaining or structuring power (typically, a few to a dozen variables are used for one type of land use change). For example, in the case of urban sprawl, typical explaining factors are the distance to existing urbanized zones or distances to roads and other means of transportation, elements of real estate costs, etc. Global scenarios are quantified in terms of global changes in land use over the whole studied area (e.g., how many hectares are

transformed from agricultural to urban uses in a given number of years, how does this evolve over time...); this is done either from academic expert knowledge, or from information provided by local planning agencies. Whenever feasible, models are validated by comparing the model predictions with actual evolution at a later date. Therefore, such models need from one to three land use maps at different dates for calibration and validation purposes (the larger the number of maps, the more robust and accurate the model). A large array of statistical tools is available in the literature to perform the calibration and validation of the model.

The horizon of projections of such models is limited in time, typically 20-30 years, due to the inherent uncertainty in such models, although they are occasionally used on longer time-scales. Climate change constraints are included, when needed, through scenarios, as it is not in the scope of such models to incorporate ecological processes that may translate climate change constraints into land cover change dynamics. Note that on such short time-scales, climate change is not dominated by the mean climate evolution but by decade variations which average out on longer time-scales and are not modeled in the global climate models used e.g. for IPCC projections for the end of the century; as a consequence, the various IPCC climate scenarios cannot be distinguished on such a short time horizon.

With regard to LUCC, the STEEP team has been involved for five years in the ESNET project whose funding came to a close in July of 2017, but the scientific production of the project is still underway. This project bears on the characterization of local Ecosystem Services networks; the project has been coordinated by LECA (*Laboratoire d'Ecologie Alpine*), in collaboration with a number of other research laboratories (most notably, IRSTEA Grenoble, besides our team), and in close interaction with a panel of local stakeholders; the scale of interest is typically a landscape (in the ecologic/geographic sense, i.e., a zone a few kilometers to a few tens of kilometers wide). The project aims at developing a generic modelling framework of ecosystem services, and studying their behavior under various scenarios of coupled urban/environment evolution, at the 2030/2040 horizon, under constraints of climate change. The contribution of the STEEP team is centered on the Land Use/Land Cover Change (LUCC) model that is one of the major building blocks of the whole project modelling effort, with the help of an ESNET funded post-doctoral researcher. In the process, areas of conceptual and methodological improvements of statistical LUCC models have been identified; implementing these improvements will be useful for the LUCC community at large, independently of the ESNET project needs.

4.3.2. *Models for Land-Use and Transportation Interactions (LUTI)*

Urban transport systems are intricately linked to urban structure and activities, i.e., to land use. Urbanization generally implies an increased travel demand. Cities have traditionally met this additional demand by extending transportation supply, through new highways and transit lines. In turn, an improvement of the accessibility of ever-farther land leads to an expansion of urban development, resulting in a significant feedback loop between transportation infrastructure and land use, one of the main causes of urban sprawl. Transportation models allow us to address questions generally limited to the impacts of new infrastructures, tolls and other legislation on traffic regulation [0], on user behavior [0], or on the environment [0]. LUTI models (Land-Use and Transport Integrated models) can answer a much broader spectrum of issues. For example, they allow us to understand how the localization of households and of economic activities (which generate transportation demand) adapt to changes of transportation supply. They also allow us to assess the impacts of such changes on the increase in real estate value, or more generally on their effects on the economic development of a specific sector or neighborhood. An economic vision interprets all these interactions in terms of equilibrium between demand and supply. Modelling the localization of households and employments (companies) relies on capturing the way stakeholders arbitrate between accessibility, real estate prices, and attractiveness of different areas.

State of the art and operability of LUTI models. The first model that proved able to analyze the interactions between transport and urbanization was developed by Lowry. Since then theories and models have become increasingly complex over time. They can be classified according to different criteria. A first classification

---

[0]Congestion, cost and time spent for the transport, etc.
[0]Changes in modality choice.
[0]CO2 emissions, air pollution, noise nuisance, etc.

retraces the historic path of these theories and models. They can be associated with one or several of the approaches underlying all present theories: economic base theory and gravity models, Input/Output models and theory of urban rent, and micro-simulations. A second possibility consists in classifying the models according to their aims and means.

Significant scientific progress has been made over the last thirty years. Nevertheless, modelling tools remain largely restricted to the academic world. Today, only seven models have at least had one recent application outside academia or are commercialized or potentially marketable, in spite of the important needs expressed by the urban planning agencies: Cube Land, DELTA, MARS, OPUS/UrbanSim, PECAS, TRANUS and Pirandello.

To guide their choice of a modelling framework, users can rely on various criteria such as the strength of the theoretical framework, the quality and the diversity of the available documentation, the accessibility of the models (is the model freely available? is the code open source? is the software regularly updated and compatible with the recent operating systems?), the functionality and friendliness of user interfaces (existence of graphic user interface, possibility of interfacing with Geographic Information Systems), existence of technical assistance, volume and availability of the data required to implement the model, etc. For example, among the seven models mentioned above, only two are open source and mature enough to meet professional standards: TRANUS and UrbanSim [0]. These two models are very different but particularly representative of the main current philosophies and trends in this scientific domain. Their comparison is informative.

STEEP implication in LUTI modelling. As yet, very few local planning authorities make use of these strategic models, mostly because they are difficult to calibrate and validate. Systematic improvement on these two critical steps would clearly increase the level of confidence in their results; these limitations hinder their dissemination in local agencies. One of the major goals of STEEP is therefore to meet the need for better calibration and validation strategies and algorithms. This research agenda lies at the core of our project CITiES (*ANR Modèles Numériques*) that ended in 2017 with the PhD defense of Thomas Capelle . This work is being partly pursued in the QAMECS project.

As for LUTI modeling, we have been using the TRANUS model since the creation of our team. In this framework we work in close collaboration with AURG [0], the local urban planning agency of Grenoble (*Agence d'Urbanisme de la Région Grenobloise*) in order to better understand and to improve the relevance of these tools for such territorial agencies.

---

[0]http://www.urbansim.org
[0]http://www.aurg.org/

TAPDANCE Team (section vide)

4. Application Domains

4.1. Controlled fusion and ITER

The search for alternative energy sources is a major issue for the future. Among others, controlled thermonuclear fusion in a hot hydrogen plasma is a promising possibility. The principle is to confine the plasma in a toroidal chamber, called a tokamak, and to attain the necessary temperatures to sustain nuclear fusion reactions. The International Thermonuclear Experimental Reactor (ITER) is a tokamak being constructed in Cadarache, France. This was the result of a joint decision by an international consortium made of the European Union, Canada, USA, Japan, Russia, South Korea, India and China. ITER is a huge project. As of today, the budget is estimated at 20 billion euros. The first plasma shot is planned for 2020 and the first deuterium-tritium operation for 2027. Many technical and conceptual difficulties have to be overcome before the actual exploitation of fusion energy. Consequently, much research has been carried out around magnetically confined fusion. Among these studies, it is important to carry out computer simulations of the burning plasma. Thus, mathematicians and computer scientists are also needed in the design of ITER. The reliability and the precision of numerical simulations allow a better understanding of the physical phenomena and thus would lead to better designs. TONUS's main involvement is in such research. The required temperatures to attain fusion are very high, of the order of a hundred million degrees. Thus it is imperative to prevent the plasma from touching the tokamak inner walls. This confinement is obtained thanks to intense magnetic fields. The magnetic field is created by poloidal coils, which generate the toroidal component of the field. The toroidal plasma current also induces a poloidal component of the magnetic field that twists the magnetic field lines. The twisting is very important for the stability of the plasma. The idea goes back to research by Tamm and Sakharov, two Russian physicists, in the 50's. Other devices are essential for the proper operation of the tokamak: divertor for collecting the escaping particles, microwave heating for reaching higher temperatures, fuel injector for sustaining the fusion reactions, toroidal coils for controlling instabilities, etc.

4.2. Other applications

The software and numerical methods that we develop can also be applied to other fields of physics or of engineering.

- For instance, we have a collaboration with the company AxesSim in Strasbourg for the development of efficient Discontinuous Galerkin (DG) solvers on hybrid computers. The applications is electromagnetic simulations for the conception of antennas, electronic devices or aircraft electromagnetic compatibility.

- The acoustic conception of large rooms requires huge numerical simulations. It is not always possible to solve the full wave equation and many reduced acoustic models have been developed. A popular model consists in considering "acoustic" particles moving at the speed of sound. The resulting Partial Differential Equation (PDE) is very similar to the Vlasov equation. The same modelling is used in radiation theory. We have started to work on the reduction of the acoustic particles model and realized that our reduction approach perfectly applies to this situation. A new PhD with CEREMA (Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement) has started in October 2015 (PhD of Pierre Gerhard). The objective is to investigate the model reduction and to implement the resulting acoustic model in our DG solver.

- In september 2017, we started a collaboration with EDF Chatou (PhD of Lucie Quibel) on the modelling of multiphase fluids with complex equations of state. The goal is to simulate the high temperature liquid-vapor flow occurring in a nuclear plant. Among others, we will apply our recent kinetic method for designing efficient implicit schemes for this kind of flows.

VIRTUAL PLANTS Project-Team (section vide)

<p style="text-align:center">**VISAGES Project-Team**</p>

4. Application Domains

4.1. Neuroimaging

One research objective in neuroimaging is the construction of anatomical and functional cerebral maps under normal and pathological conditions. Many researches are currently performed to find correlations between anatomical structures, essentially sulci and gyri, where neuronal activation takes place, and cerebral functions, as assessed by recordings obtained by the means of various neuroimaging modalities, such as PET (Positron Emission Tomography), fMRI (Functional Magnetic Resonance Imaging), EEG (Electro-EncephaloGraphy) and MEG (Magneto-EncephaloGraphy). Then, a central problem inherent to the formation of such maps is to put together recordings obtained from different modalities and from different subjects. This mapping can be greatly facilitated by the use of MR anatomical brain scans with high spatial resolution that allows a proper visualization of fine anatomical structures (sulci and gyri). Recent improvements in image processing techniques, such as segmentation, registration, delineation of the cortical ribbon, modeling of anatomical structures and multi-modality fusion, make possible this ambitious goal in neuroimaging. This problem is very rich in terms of applications since both clinical and neuroscience applications share similar problems. Since this domain is very generic by nature, our major contributions are directed towards clinical needs even though our work can address some specific aspects related to the neuroscience domain.

4.2. Multiple sclerosis

Over the past years, a discrepancy became apparent between clinical Multiple sclerosis (MS) classification describing on the one hand MS according to four different disease courses and, on the other hand, the description of two different disease stages (an early inflammatory and a subsequently neurodegenerative phase). It is to be expected that neuroimaging will play a critical role to define in vivo those four different MS lesion patterns. An in vivo distinction between the four MS lesion patterns, and also between early and late stages of MS will have an important impact in the future for a better understanding of the natural history of MS and even more for the appropriate selection and monitoring of drug treatment in MS patients. MRI has a low specificity for defining in more detail the pathological changes which could discriminate between the different lesion types. However, it has a high sensitivity to detect focal and also widespread, diffuse pathology of the normal appearing white and gray matter. Our major objective within this application domain is then to define new neuroimaging markers for tracking the evolution of the pathology from high dimensional data (e.g., nD+t MRI) in the brain and the spinal cord. In addition, in order to complement MR neuroimaging data, we ambition to perform also cell labeling neuroimaging (e.g., MRI or PET) and to compare MR and PET data using standard and experimental MR contrast agents and radiolabeled PET tracers for activated microglia (e.g., USPIO or PK 11195). The goal is to define and develop, for routine purposes, cell specific and also quantitative imaging markers for the improved in vivo characterization of MS pathology.

4.3. Modeling of anatomical and anatomo-functional neurological patterns

The major objective within this application domain is to build anatomical and functional brain atlases in the context of functional mapping and for the study of developmental, neurodegenerative or even psychiatric brain diseases (Multiple sclerosis, Epilepsy, Parkinson, Dysphasia, Depression or even Alzheimer). This is a very competitive research domain; our contribution is based on our previous works in this field, and by continuing our local and wider collaborations. An additional objective within this application domain is to find new descriptors to study the brain anatomy and/or function (e.g., variation of brain perfusion, evolution in shape and size of an anatomical structure in relation with pathology or functional patterns, computation of asymmetries ...). This is also a very critical research domain, especially for many developmental or neurodegenerative brain diseases.

XPOP Project-Team

4. Application Domains

4.1. Precision medicine and pharmacogenomics

Pharmacogenomics involves using an individual's genome to determine whether or not a particular therapy, or dose of therapy, will be effective. Indeed, people's reaction to a given drug depends on their physiological state and environmental factors, but also to their individual genetic make-up.

Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person. While some advances in precision medicine have been made, the practice is not currently in use for most diseases.

Currently, in the traditional population approach, inter-individual variability in the reaction to drugs is modeled using covariates such as weight, age, sex, ethnic origin, etc. Genetic polymorphisms susceptible to modify pharmacokinetic or pharmacodynamic parameters are much harder to include, especially as there are millions of possible polymorphisms (and thus covariates) per patient.

The challenge is to determine which genetic covariates are associated to some PKPD parameters and/or implicated in patient responses to a given drug.

Another problem encountered is the dependence of genes, as indeed, gene expression is a highly regulated process. In cases where the explanatory variables (genomic variants) are correlated, Lasso-type methods for model selection are thwarted.

There is therefore a clear need for new methods and algorithms for the estimation, validation and selection of mixed effects models adapted to the problems of genomic medicine.

A target application of this project concerns the lung cancer.

EGFR (Epidermal Growth Factor Receptor) is a cell surface protein that binds to epidermal growth factor. We know that deregulation of the downstream signaling pathway of EGFR is involved in the development of lung cancers and several gene mutations responsible for this deregulation are known.

Our objective is to identify the variants responsible for the disruption of this pathway using a modelling approach. The data that should be available for developing such model are ERK (Extracellular signal–regulated kinases) phosphorylation time series, obtained from different genetic profiles.

The model that we aim to develop will describe the relationship between the parameters of the pathway and the genomic covariates, i.e. the genetic profile. Variants related to the pathway include: variants that modify the affinity binding of ligands to receptors, variants that modify the total amount of protein, variants that affect the catalytic site,...

4.2. Oncology

In cancer, the most dreadful event is the formation of metastases that disseminate tumor cells throughout the organism. Cutaneous melanoma is a cancer, where the primary tumor can easily be removed by surgery. However, this cancer is of poor prognosis; because melanomas metastasize often and rapidly. Many melanomas arise from excessive exposure to mutagenic UV from the sun or sunbeds. As a consequence, the mutational burden of melanomas is generally high

RAC1 encodes a small GTPase that induces cell cycle progression and migration of melanoblasts during embryonic development. Patients with the recurrent P29S mutation of RAC1 have 3-fold increased odds at having regional lymph nodes invaded at the time of diagnosis. RAC1 is unlikely to be a good therapeutic target, since a potential inhibitor that would block its catalytic activity, would also lock it into the active GTP-bound state. This project thus investigates the possibility of targeting the signaling pathway downstream of RAC1.

XPOP is mainly involved in Task 1 of the project: *Identifying deregulations and mutations of the ARP2/3 pathway in melanoma patients.*

Association of over-expression or down-regulation of each marker with poor prognosis in terms of invasion of regional lymph nodes, metastases and survival, will be examined using classical univariate and multivariate analysis. We will then develop specific statistical models for survival analysis in order to associate prognosis factors to each composition of complexes. Indeed, one has to implement the further constraint that each subunit has to be contributed by one of several paralogous subunits. An original method previously developed by XPOP has already been successfully applied to WAVE complex data in breast cancer.

The developed models will be rendered user-friendly though a dedicated Rsoftware package.

This project can represent a significant step forward in precision medicine of the cutaneous melanoma.

4.3. Hemodialysis

Hemodialysis is a process for removing waste and excess water from the blood and is used primarily as an artificial replacement for lost kidney function in people with kidney failure. Side effects caused by removing too much fluid and/or removing fluid too rapidly include low blood pressure, fatigue, chest pains, leg-cramps, nausea and headaches.

Nephrologists must therefore correctly assess the hydration status in chronic hemodialysis patients and consider fluid overload effects when prescribing dialysis, according to a new study.

The fluid overload biomarker, B-type natriuretic peptide (BNP) is an important component of managing patients with kidney disease. Indeed, it is believed that each dialysis patient will have an ideal or "dry" BNP level which will accurately and reproducibly reflect their optimal fluid status.

The objective of this study is to develop a model for the BNP and the hydratation status using individual information (age, sex, ethnicity, systolic blood pressure, BMI, coronary heart disease history, ...).

The impact will be significant if the method succeeds. Indeed, it will be possible for the nephrologists to use this model for monitoring individually each treatment, in order to avoid risks of hypotension (low BNP) or overweight (high BNP).

4.4. Intracellular processes

Significant cell-to-cell heterogeneity is ubiquitously-observed in isogenic cell populations. Cells respond differently to a same stimulation. For example, accounting for such heterogeneity is essential to quantitatively understand why some bacteria survive antibiotic treatments, some cancer cells escape drug-induced suicide, stem cell do not differentiate, or some cells are not infected by pathogens.

The origins of the variability of biological processes and phenotypes are multifarious. Indeed, the observed heterogeneity of cell responses to a common stimulus can originate from differences in cell phenotypes (age, cell size, ribosome and transcription factor concentrations, etc), from spatio-temporal variations of the cell environments and from the intrinsic randomness of biochemical reactions. From systems and synthetic biology perspectives, understanding the exact contributions of these different sources of heterogeneity on the variability of cell responses is a central question.

The main ambition of this project is to propose a paradigm change in the quantitative modelling of cellular processes by shifting from mean-cell models to single-cell and population models. The main contribution of XPOP focuses on methodological developments for mixed-effects model identification in the context of growing cell populations.

- Mixed-effects models usually consider an homogeneous population of independent individuals. This assumption does not hold when the population of cells (i.e. the statistical individuals) consists of several generations of dividing cells. We then need to account for inheritance of single-cell parameters in this population. More precisely, the problem is to attribute the new state and parameter values to newborn cells given (the current estimated values for) the mother.

- The mixed-effects modelling framework corresponds to a strong assumption: differences between cells are static in time (ie, cell-specific parameters have fixed values). However, it is likely that for any given cell, ribosome levels slowly vary across time, since like any other protein, ribosomes are produced in a stochastic manner. We will therefore extend our modelling framework so as to account for the possible random fluctuations of parameter values in individual cells. Extensions based on stochastic differential equations will be investigated.

- Identifiability is a fundamental prerequisite for model identification and is also closely connected to optimal experimental design. We will derive criteria for theoretical identifiability, in which different parameter values lead to non-identical probability distributions, and for structural identifiability, which concerns the algebraic properties of the structural model, i.e. the ODE system. We will then address the problem of practical identifiability, whereby the model may be theoretically identifiable but the design of the experiment may make parameter estimation difficult and imprecise. An interesting problem is whether accounting for lineage effects can help practical identifiability of the parameters of the individuals in presence of measurement and biological noise.

4.5. Population pharmacometrics

Pharmacometrics involves the analysis and interpretation of data produced in pre-clinical and clinical trials. Population pharmacokinetics studies the variability in drug exposure for clinically safe and effective doses by focusing on identification of patient characteristics which significantly affect or are highly correlated with this variability. Disease progress modeling uses mathematical models to describe, explain, investigate and predict the changes in disease status as a function of time. A disease progress model incorporates functions describing natural disease progression and drug action.

The model based drug development (MBDD) approach establishes quantitative targets for each development step and optimizes the design of each study to meet the target. Optimizing study design requires simulations, which in turn require models. In order to arrive at a meaningful design, mechanisms need to be understood and correctly represented in the mathematical model. Furthermore, the model has to be predictive for future studies. This requirement precludes all purely empirical modeling; instead, models have to be mechanistic.

In particular, physiologically based pharmacokinetic models attempt to mathematically transcribe anatomical, physiological, physical, and chemical descriptions of phenomena involved in the ADME (Absorption - Distribution - Metabolism - Elimination) processes. A system of ordinary differential equations for the quantity of substance in each compartment involves parameters representing blood flow, pulmonary ventilation rate, organ volume, etc.

The ability to describe variability in pharmacometrics model is essential. The nonlinear mixed-effects modeling approach does this by combining the structural model component (the ODE system) with a statistical model, describing the distribution of the parameters between subjects and within subjects, as well as quantifying the unexplained or residual variability within subjects.

The objective of XPOP is to develop new methods for models defined by a very large ODE system, a large number of parameters and a large number of covariates. Contributions of XPOP in this domain are mainly methodological and there is no privileged therapeutic application at this stage.

However, it is expected that these new methods will be implemented in software tools, including MONOLIX and Rpackages for practical use.

AGORA Team

4. Application Domains

4.1. Smart Cities

One major characteristic of modern societies is that they are prevalently urban. In coherence, the contributions of the Agora team are in particular applied to provide solutions tailored to the emergence of the Internet of Things (IoT) and to Smart Cities applications. A major motivation of the team is the forthcoming explosion of the number of connected devices. In particular, low cost - small data devices are supposed to be densely deployed in our environment, fostering the interest for a convergence of the traditional wireless networking paradigms.

Smart City is a constantly reshaped concept, embracing the future of dense metropolitan areas, with references to efficient and sustainable infrastructure, improving citizens' quality of life and protecting the environment. A consensus on the Smart City philosophy is however that it will be primarily achieved by leveraging a clever integration of Information and Communication Technologies (ICT) in the urban tissue. Indeed, ICTs are enabling an evolution from the current duality between the real world and its digitized counterpart to a continuum in which digital contents and applications are seamlessly interacting with classical infrastructures and services. Smart Cities are often described by the digital services that should be provided which are inherently dependent on dense measurements of the city environment and activities, the collection of these data, their processing into information, and their redistribution. The networking infrastructure plays therefore a critical role in enabling advanced services, in particular the wireless infrastructure supporting density and mobility.

From a wireless networking viewpoint, the digitization of cities can be seen as a paradigm shift extending the Internet of Things (IoT) to a citizen-centric model in order to leverage the massive data collected by pervasive sensors, connected mobiles or fixed devices, and social applications.

<p style="text-align:center">**ALPINES Project-Team**</p>

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 CO_2 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 subregion 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 functional framework that naturally involves singular functions, which are difficult to 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 characterizing 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 Moore's Law. Planck (http://planck.esa.int/) 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.

ASAP Project-Team (section vide)

ASCOLA Project-Team

4. Application Domains

4.1. Enterprise Information Systems and Services

Large IT infrastructures typically evolve by adding new third-party or internally-developed components, but also frequently by integrating already existing information systems. Integration frequently requires the addition of glue code that mediates between different software components and infrastructures but may also consist in more invasive modifications to implementations, in particular to implement crosscutting functionalities. In more abstract terms, enterprise information systems are subject to structuring problems involving horizontal composition (composition of top-level functionalities) as well as vertical composition (reuse and sharing of implementations among several top-level functionalities). Moreover, information systems have to be more and more dynamic.

Service-Oriented Computing (SOC) that is frequently used for solving some of the integration problems discussed above. Indeed, service-oriented computing has two main advantages:

- Loose-coupling: services are autonomous: they do not require other services to be executed;
- Ease of integration: Services communicate over standard protocols.

Our current work is based on the following observation: similar to other compositional structuring mechanisms, SOAs are subject to the problem of crosscutting functionalities, that is, functionalities that are scattered and tangled over large parts of the architecture and the underlying implementation. Security functionalities, such as access control and monitoring for intrusion detection, are a prime example of such a functionality in that it is not possible to modularize security issues in a well-separated module. Aspect-Oriented Software Development is precisely an application-structuring method that addresses in a systemic way the problem of the lack of modularization facilities for crosscutting functionalities.

We are considering solutions to secure SOAs by providing an aspect-oriented structuring and programming model that allows security functionalities to be modularized. Two levels of research have been identified:

- Service level: as services can be composed to build processes, aspect weaving will deal with the orchestration and the choreography of services.
- Implementation level: as services are abstractly specified, aspect weaving will require to extend service interfaces in order to describe the effects of the executed services on the sensitive resources they control.

4.2. Capacity Planning in Cloud, Fog and Edge Computing

Cloud and more recently Fog and Edge computing platforms aim at delivering large capacities of computing power. These capacities can be used to improve performance (for scientific applications) or availability (e.g., for Internet services hosted by datacenters). These distributed infrastructures consist of a group of coupled computers that work together and may be spread across a LAN (cluster), across a the Internet (Fog/Edge). Due to their large scale, these architectures require permanent adaptation, from the application to the system level and call for automation of the corresponding adaptation processes. We focus on self-configuration and self-optimization functionalities across the whole software stack: from the lower levels (systems mechanisms such as distributed file systems for instance) to the higher ones (i.e. the applications themselves such as clustered servers or scientific applications).

In 2017, we have been consolidating our expertise around the OpenStack ecosystem. We proposed in particular EnOS, a dedicated framework to conduct performance analyses of OpenStack at large-scale ina reproducible manner. The framework enables researchers to conduct experiments in an automize manner on top of different testbeds such as Grid'5000 and Chameleon. see Sec. 7.1 .

In the energy field, we have designed a set of techniques, named Optiplace, for cloud management with flexible power models through constraint programming. OptiPlace supports external models, named views. Specifically, we have developed a power view, based on generic server models, to define and reduce the power consumption of a datacenter's physical servers. We have shown that OptiPlace behaves at least as good as our previous system, Entropy, requiring as low as half the time to find a solution for the constrained-based placement of tasks for large datacenters.

4.3. Pervasive Systems

Pervasive systems are another class of systems raising interesting challenges in terms of software structuring. Such systems are highly concurrent and distributed. Moreover, they assume a high-level of mobility and context-aware interactions between numerous and heterogeneous devices (laptops, PDAs, smartphones, cameras, electronic appliances...). Programming such systems requires proper support for handling various interfering concerns like software customization and evolution, security, privacy, context-awareness... Additionally, service composition occurs spontaneously at runtime.

Like Pervasive systems, Internet of thing is a major theme of these last ten years. Many research works has been led on the whole chain, from communicating sensors to big data management, through communication middlewares.

The more a sensor networks senses various data, the more the users panel is heterogeneous. Such an heterogeneity leads to a major problem about data modeling: for each user, to aim at precisely addressing his needs and his needs only; ie to avoid a data representation which would overwhelm the user with all the data sensed from the network, regardless if he needs it or not. To leverage this issue, we propose a multitree modeling for sensor networks which addresses each of these specific usages.With this modeling comes a domain specific language (DSL) which allows users to manipulate, parse and aggregate information from the sensors.

<p align="center" style="color:red">AVALON Project-Team</p>

4. Application Domains

4.1. Overview

The Avalon team targets applications with large computing and/or data storage needs, which are still difficult to program, maintain, and deploy. Those applications can be parallel and/or distributed applications, such as large scale simulation applications or code coupling applications. Applications can also be workflow-based as commonly found in distributed systems such as grids or clouds.

The team aims at not being restricted to a particular application field, thus avoiding any spotlight. The team targets different HPC and distributed application fields, which bring use cases with different issues. This will be eased by our various collaborations: the team participates to the INRIA-Illinois Joint Laboratory for Petascale Computing, the Physics, Radiobiology, Medical Imaging, and Simulation French laboratory of excellence, the E-Biothon project, the INRIA large scale initiative Computer and Computational Sciences at Exascale (C2S@Exa), and to BioSyL, a federative research structure about Systems Biology of the University of Lyon. Moreover, the team members have a long tradition of cooperation with application developers such as CERFACS and EDF R&D. Last but not least, the team has a privileged connection with CC IN2P3 that opens up collaborations, in particular in the astrophysics field.

In the following, some examples of representative applications we are targeting are presented. In addition to highlighting some application needs, they also constitute some of the use cases we will use to valide our theoretical results.

4.2. Climatology

The world's climate is currently changing due to the increase of the greenhouse gases in the atmosphere. Climate fluctuations are forecasted for the years to come. For a proper study of the incoming changes, numerical simulations are needed, using general circulation models of a climate system. Simulations can be of different types: HPC applications (*e.g.,* the NEMO framework [38] for ocean modelization), code-coupling applications (*e.g.,* the OASIS coupler [44] for global climate modeling), or workflows (long term global climate modeling).

As for most applications the team is targeting, the challenge is to thoroughly analyze climate-forecasting applications to model their needs in terms of programing model, execution model, energy consumption, data access pattern, and computing needs. Once a proper model of an application has been set up, appropriate scheduling heuristics could be designed, tested, and compared. The team has a long tradition of working with CERFACS on this topic, for example in the LEGO (2006-09) and SPADES (2009-12) French ANR projects.

4.3. Astrophysics

Astrophysics is a major field to produce large volumes of data. For instance, the Large Synoptic Survey Telescope (http://www.lsst.org/lsst/) will produce 15 TB of data every night, with the goals of discovering thousands of exoplanets and of uncovering the nature of dark matter and dark energy in the universe. The Square Kilometer Array (http://www.skatelescope.org/) produces 9 Tbits/s of raw data. One of the scientific projects related to this instrument called Evolutionary Map of the Universe is working on more than 100 TB of images. The Euclid Imaging Consortium (https://www.euclid-ec.org/) will generate 1 PB data per year.

Avalon collaborates with the *Institut de Physique Nucléaire de Lyon* (IPNL) laboratory on large scale numerical simulations in astronomy and astrophysics. Contributions of the Avalon members have been related to algorithmic skeletons to demonstrate large scale connectivity, the development of procedures for the generation of realistic mock catalogs, and the development of a web interface to launch large cosmological simulations on GRID'5000.

This collaboration, that continues around the topics addressed by the CLUES project (http://www.clues-project.org), has been extended thanks to the tight links with the CC-IN2P3. Major astrophysics projects execute part of their computing, and store part of their data on the resources provided by the CC-IN2P3. Among them, we can mention SNFactory, Euclid, or LSST. These applications constitute typical use cases for the research developed in the Avalon team: they are generally structured as workflows and a huge amount of data (from TB to PB) is involved.

4.4. Bioinformatics

Large-scale data management is certainly one of the most important applications of distributed systems in the future. Bioinformatics is a field producing such kinds of applications. For example, DNA sequencing applications make use of MapReduce skeletons.

The Avalon team is a member of BioSyL (http://www.biosyl.org), a Federative Research Structure attached to University of Lyon. It gathers about 50 local research teams working on systems biology. Moreover, the team cooperated with the French Institute of Biology and Chemistry of Proteins (IBCP http://www.ibcp.fr) in particular through the ANR MapReduce project where the team focuses on a bio-chemistry application dealing with protein structure analysis. Avalon has also started working with the Inria Beagle team (https://team.inria.fr/beagle/) on artificial evolution and computational biology as the challenges are around high performance computation and data management.

COAST Project-Team (section vide)

<div align="center">

COATI Project-Team

</div>

4. Application Domains

4.1. Telecommunication Networks

COATI is mostly interested in telecommunications networks. We focus on the design and management of heterogeneous physical and logical networks. The project has kept working on the design of backbone networks (optical networks, radio networks, IP networks). We also study routing algorithms such as dynamic and compact routing schemes, as we did in the context of the FP7 EULER led by Alcatel-Lucent Bell-Labs (Belgium), and the evolution of the routing in case of any kind of topological modifications (maintenance operations, failures, capacity variations, etc.). However, the fields of Software Defined Networks and Network Function Virtualization are growing in importance in our studies.

4.2. Other Domains

Our combinatorial tools may be well applied to solve many other problems in various areas (transport, biology, resource allocation, chemistry, smart-grids, speleology, etc.) and we intend to collaborate with experts of these other domains.

For instance, we collaborate with EP ABS (Algorithms Biology Structure) from Sophia Antipolis on problems from Structural Biology (see Section 7.2.1.5). In the area of transportation networks, we have started a collaboration with SME Instant-System on dynamic car-pooling combined with multi-modal transportation systems. This collaboration will be strengthen in the near future with the support of an ANR project (starting January 2018).

<p style="text-align:center;color:red;">CTRL-A Project-Team</p>

4. Application Domains

4.1. Self-adaptive and reconfigurable computing systems in HPC and the IoT

We are attacking the problem of designing well-regulated and efficient self-adaptive computing systems by the development of novel strategies for systems management.

The kind of systems we typically target involve relatively coarse grained computation tasks (e.g. image processing or HPC tasks, components or services), assembled in workflows, application dependency graphs, or composites. At that level, there can be parallel and conditional branches, as well as choices that can be made between alternative branches, corresponding to different ways to perform that part of the application. Such tasks can be achieved following a choice of implementations or versions, such as in service oriented approaches. Each implementation has its own characteristics and requirements, e.g., w.r.t. resources consumed and QoS offered. The systems execution infrastructures present heterogeneity, with different computing processors, a variety of peripheral devices (e.g., I/O, video port, accelerators), and different means of communications. This hardware or middleware level also presents adaptation potential e.g., in varying quantities of resources or sleep and stand-by modes.

The kinds of control problems encountered in these adaptive systems concern the navigation in the configurations space defined by choice points at the levels of applications, tasks, and architecture. Upon events or conditions triggering reconfiguration and adaptation, the controller has to choose a next configuration where, on the one hand, all consistency constraints are satisfied w.r.t. dependencies and resources requirements. On the other hand, it has to apply a policy or strategy deciding between correct configurations which one to chose e.g. by optimizing one or more criteria, or by insuring reachability of some later configuration (goal or fallback). This targeted class of computing systems we consider is mid-sized, in the sense that the combinatorial complexity is large enough for manual solving to be impossible, while remaining within the range where supervisory control techniques are tractable. The pace of control is more sporadic, and slower than the instruction-level computation performance within the coarse-grained tasks.

The objectives of CTRL-A will be achieved and evaluated in both of our main application domains, thereby exhibiting their similarities from the point of view of reconfiguration control. A first application domain is High Performance Computing. In this area, we currently focus especially on the management of Dynamic Partial Reconfiguration in FPGA based hardware, at the level of middleware. Here the particular application we consider is, as in our ANR project HPeC starting end of 2015, video image flow processing for smart cameras implemented on DPR FPGASs themselves embedded in drones.

A second general application domain to confront our models is the Internet of Things (IoT), more specifically self-adaptive middleware platforms for Smart Environments, or Industry 4.0 related topics, like SCADA. We focus on providing coordination components and controllers of software components and services, or rule-based middleware platforms. The adaptation problems concern both the functional aspects of the applications in a smart building, and the middleware support deployment and reconfiguration issues. We are considering perspectives concerning self-protection and security.

DANTE Project-Team

4. Application Domains

4.1. Life Science & Health

In parallel to the advances in modern medicine, health sciences and public health policy, epidemic models aided by computer simulations and information technologies offer an increasingly important tool for the understanding of transmission dynamics and of epidemic patterns. The increased computational power and use of Information and Communication Technologies make feasible sophisticated modelling approaches augmented by detailed in vivo data sets, and allow to study a variety of possible scenarios and control strategies, helping and supporting the decision process at the scientific, medical and public health level. The research conducted in the DANTE project finds direct applications in the domain of LSH since modelling approaches crucially depend on our ability to describe the interactions of individuals in the population. In the MOSAR/iBird project we are collaborating with the team of Pr. Didier Guillemot (Inserm/Institut. Pasteur/Université de Versailles). Within the TUBEXPO and ARIBO projects, we are collaborating with Pr. Jean-Christopge Lucet (Professeur des université Paris VII, Praticien hospitalier APHP).

4.2. Network Science / Complex networks

In the last ten years the science of complex networks has been assigned an increasingly relevant role in defining a conceptual framework for the analysis of complex systems. Network science is concerned with graphs that map entities and their interactions to nodes and links. For a long time, this mathematical abstraction has contributed to the understanding of real-world systems in physics, computer science, biology, chemistry, social sciences, and economics. Recently, however, enormous amounts of detailed data, electronically collected and meticulously catalogued, have finally become available for scientific analysis and study. This has led to the discovery that most networks describing real world systems show the presence of complex properties and heterogeneities, which cannot be neglected in their topological and dynamical description. This has called forth a major effort in developing the methodology to characterise the topology and temporal behaviour of complex networks, to describe the observed structural and temporal heterogeneities, to detect and measure emerging community structure, to see how the functionality of networks determines their evolving structure, and to determine what kinds of correlations play a role in their dynamics. All these efforts have brought us to a point where the science of complex networks has become advanced enough to help us to disclose the deeper roles of complexity and gain understanding about the behaviour of very complicated systems.

In this endeavour the DANTE project targets the study of dynamically evolving networks, concentrating on questions about the evolving structure and dynamical processes taking place on them. During the last year we developed developed several projects along these lines concerning three major datasets:

- Mobile telephony data: In projects with academic partners and Grandata we performed projects based on two large independent datasets collecting the telephone call and SMS event records for million of anonymised individuals. The datasets record the time and duration of mobile phone interactions and some coarse grained location and demographic data for some users. In addition one of the dataset is coupled with anonymised bank credit information allowing us to study directly the socioeconomic structure of a society and how it determines the communication dynamics and structure of individuals.

- Skype data: Together with Skype Labs/STACC and other academic groups we were leading projects in the subject of social spreading phenomena. These projects were based on observations taken from a temporally detailed description of the evolving social network of (anonymised) Skype users registered between 2003 and 2011. This data contains dates of registration and link creation together with gradual information about their location and service usage dynamics.

- Twitter data: In collaboration with ICAR-ENS Lyon we collected a large dataset about the microblogs and communications of millions of Twitter users in the French Twitter space. This data allows us to follow the spreading of fads/opinions/hashtags/ideas and more importantly linguistic features in online communities. The aim of this collaboration is to set the ground for a quantitative framework studying the evolution of linguistic features and dialects in an social-communication space mediated by online social interactions.

DATAMOVE Project-Team

4. Application Domains

4.1. Data Aware Batch Scheduling

Large scale high performance computing platforms are becoming increasingly complex. Determining efficient allocation and scheduling strategies that can adapt to technological evolutions is a strategic and difficult challenge. We are interested in scheduling jobs in hierarchical and heterogeneous large scale platforms. On such platforms, application developers typically submit their jobs in centralized waiting queues. The job management system aims at determining a suitable allocation for the jobs, which all compete against each other for the available computing resources. Performances are measured using different classical metrics like maximum completion time or slowdown. Current systems make use of very simple (but fast) algorithms that however rely on simplistic platform and execution models, and thus, have limited performances.

For all target scheduling problems we aim to provide both theoretical analysis and complementary analysis through simulations. Achieving meaningful results will require strong improvements on existing models (on power for example) and the design of new approximation algorithms with various objectives such as stretch, reliability, throughput or energy consumption, while keeping in focus the need for a low-degree polynomial complexity.

4.1.1. Algorithms

The most common batch scheduling policy is to consider the jobs according to the First Come First Served order (FCFS) with backfilling (BF). BF is the most widely used policy due to its easy and robust implementation and known benefits such as high system utilization. It is well-known that this strategy does not optimize any sophisticated function, but it is simple to implement and it guarantees that there is no starvation (i.e. every job will be scheduled at some moment).

More advanced algorithms are seldom used on production platforms due to both the gap between theoretical models and practical systems and speed constraints. When looking at theoretical scheduling problems, the generally accepted goal is to provide polynomial algorithms (in the number of submitted jobs and the number of involved computing units). However, with millions of processing cores where every process and data transfer have to be individually scheduled, polynomial algorithms are prohibitive as soon as the polynomial degree is too large. The model of *parallel tasks* simplifies this problem by bundling many threads and communications into single boxes, either rigid, rectangular or malleable. Especially malleable tasks capture the dynamicity of the execution. Yet these models are ill-adapted to heterogeneous platforms, as the running time depends on more than simply the number of allotted resources, and some of the common underlying assumptions on the speed-up functions (such as monotony or concavity) are most often only partially verified.

In practice, the job execution times depend on their allocation (due to communication interferences and heterogeneity in both computation and communication), while theoretical models of parallel jobs usually consider jobs as black boxes with a fixed (maximum) execution time. Though interesting and powerful, the classical models (namely, synchronous PRAM model, delay, LogP) and their variants (such as hierarchical delay), are not well-suited to large scale parallelism on platforms where the cost of moving data is significant, non uniform and may change over time. Recent studies are still refining such models in order to take into account communication contentions more accurately while remaining tractable enough to provide a useful tool for algorithm design.

Today, all algorithms in use in production systems are oblivious to communications. One of our main goals is to **design a new generation of scheduling algorithms fitting more closely job schedules according to platform topologies**.

4.1.2. Locality Aware Allocations

Recently, we developed modifications of the standard back-filling algorithm taking into account platform topologies. The proposed algorithms take into account locality and contiguity in order to hide communication patterns within parallel tasks. The main result here is to establish good lower bounds and small approximation ratios for policies respecting the locality constraints. The algorithms work in an online fashion, improving the global behavior of the system while still keeping a low running time. These improvements rely mainly on our past experience in designing approximation algorithms. Instead of relying on complex networking models and communication patterns for estimating execution times, the communications are disconnected from the execution time. Then, the scheduling problem leads to a trade-off: optimizing locality of communications on one side and a performance objective (like the makespan or stretch) on the other side.

In the perspective of taking care of locality, other ongoing works include the study of schedulers for platforms whose interconnection network is a static structured topology (like the 3D-torus of the BlueWaters platform we work on in collaboration with the Argonne National Laboratory). One main characteristic of this 3D-torus platform is to provide I/O nodes at specific locations in the topology. Applications generate and access specific data and are thus bounded to specific I/O nodes. Resource allocations are constrained in a strong and unusual way. This problem is close for actual hierarchical platforms. The scheduler needs to compute a schedule such that I/O nodes requirements are filled for each application while at the same time avoiding communication interferences. Moreover, extra constraints can arise for applications requiring accelerators that are gathered on the nodes at the edge of the network topology.

While current results are encouraging, they are however limited in performance by the low amount of information available to the scheduler. We look forward to extend ongoing work by progressively increasing application and network knowledge (by technical mechanisms like profiling or monitoring or by more sophisticated methods like learning). It is also important to anticipate on application resource usage in terms of compute units, memory as well as network and I/Os to efficiently schedule a mix of applications with different profiles. For instance, a simple solution is to partition the jobs as "communication intensive" or "low communications". Such a tag could be achieved by the users them selves or obtained by learning techniques. We could then schedule low communications jobs using leftover spaces while taking care of high communication jobs. More sophisticated options are possible, for instance those that use more detailed communication patterns and networking models. Such options would leverage the work proposed in Section 4.2 for gathering application traces.

4.1.3. Data-Centric Processing

Exascale computing is shifting away from the traditional compute-centric models to a more data-centric one. This is driven by the evolving nature of large scale distributed computing, no longer dominated by pure computations but also by the need to handle and analyze large volumes of data. These data can be large databases of results, data streamed from a running application or another scientific instrument (collider for instance). These new workloads call for specific resource allocation strategies.

Data movements and storage are expected to be a major energy and performance bottleneck on next generation platforms. Storage architectures are also evolving, the standard centralized parallel file system being complemented with local persistent storage (Burst Buffers, NVRAM). Thus, one data producer can stage data on some nodes' local storage, requiring to schedule close by the associated analytics tasks to limit data movements. This kind of configuration, often referred as *in-situ analytics*, is expected to become common as it enables to switch from the traditional I/O intensive workflow (batch-processing followed by *post mortem* analysis and visualization) to a more storage conscious approach where data are processed as closely as possible to where and when they are produced (in-situ processing is addressed in details in section 4.3). By reducing data movements and scheduling the extra processing on resources not fully exploited yet, in-situ processing is expected to have also a significant positive energetic impact. Analytics codes can be executed in the same nodes than the application, often on dedicated cores commonly called helper cores, or on dedicated nodes called staging nodes. The results are either forwarded to the users for visualization or saved to disk through I/O nodes. In-situ analytics can also take benefit of node local disks or burst buffers to reduce data movements.

Future job scheduling strategies should take into account in-situ processes in addition to the job allocation to optimize both energy consumption and execution time. On the one hand, this problem can be reduced to an allocation problem of extra asynchronous tasks to idle computing units. But on the other hand, embedding analytics in applications brings extra difficulties by making the application more heterogeneous and imposing more constraints (data affinity) on the required resources. Thus, the main point here is to develop efficient algorithms for dealing with heterogeneity without increasing the global computational cost.

4.1.4. Learning

Another important issue is to adapt the job management system to deal with the bad effects of uncertainties, which may be catastrophic in large scale heterogeneous HPC platforms (jobs delayed arbitrarly far or jobs killed). A natural question is then: *is it possible to have a good estimation of the job and platform parameters in order to be able to obtain a better scheduling ?* Many important parameters (like the number or type of required resources or the estimated running time of the jobs) are asked to the users when they submit their jobs. However, some of these values are not accurate and in many cases, they are not even provided by the end-users. In DataMove, we propose to study new methods for a better prediction of the characteristics of the jobs and their execution in order to improve the optimization process. In particular, the methods well-studied in the field of big data (in supervised Machine Learning, like classical regression methods, Support Vector Methods, random forests, learning to rank techniques or deep learning) could and must be used to improve job scheduling in large scale HPC platforms. This topic received a great attention recently in the field of parallel and distributed processing. A preliminary study has been done recently by our team with the target of predicting the job running times (called wall times). We succeeded to improve significantly in average the reference EASY Back Filling algorithm by estimating the wall time of the jobs, however, this method leads to big delay for the stretch of few jobs. Even if we succeed in determining more precisely hidden parameters, like the wall time of the jobs, this is not enough to determine an optimized solution. The shift is not only to learn on dedicated parameters but also on the scheduling policy. The data collected from the accounting and profiling of jobs can be used to better understand the needs of the jobs and through learning to propose adaptations for future submissions. The goal is to propose extensions to further improve the job scheduling and improve the performance and energy efficiency of the application. For instance preference learning may enable to compute on-line new priorities to back-fill the ready jobs.

4.1.5. Multi-objective Optimization

Several optimization questions that arise in allocation and scheduling problems lead to the study of several objectives at the same time. The goal is then not a single optimal solution, but a more complicated mathematical object that captures the notion of trade-off. In broader terms, the goal of multi-objective optimization is not to externally arbitrate on disputes between entities with different goals, but rather to explore the possible solutions to highlight the whole range of interesting compromises. A classical tool for studying such multi-objective optimization problems is to use *Pareto curves*. However, the full description of the Pareto curve can be very hard because of both the number of solutions and the hardness of computing each point. Addressing this problem will opens new methodologies for the analysis of algorithms.

To further illustrate this point here are three possible case studies with emphasis on conflicting interests measured with different objectives. While these cases are good representatives of our HPC context, there are other pertinent trade-offs we may investigate depending on the technology evolution in the coming years. This enumeration is certainly not limitative.

Energy versus Performance. The classical scheduling algorithms designed for the purpose of performance can no longer be used because performance and energy are contradictory objectives to some extent. The scheduling problem with energy becomes a multi-objective problem in nature since the energy consumption should be considered as equally important as performance at exascale. A global constraint on energy could be a first idea for determining trade-offs but the knowledge of the Pareto set (or an approximation of it) is also very useful.

Administrators versus application developers. Both are naturally interested in different objectives: In current algorithms, the performance is mainly computed from the point of view of administrators, but the users should be in the loop since they can give useful information and help to the construction of better schedules. Hence, we face again a multi-objective problem where, as in the above case, the approximation of the Pareto set provides the trade-off between the administrator view and user demands. Moreover, the objectives are usually of the same nature. For example, *max stretch* and *average stretch* are two objectives based on the slowdown factor that can interest administrators and users, respectively. In this case the study of the norm of stretch can be also used to describe the trade-off (recall that the L_1-norm corresponds to the average objective while the L_∞-norm to the max objective). Ideally, we would like to design an algorithm that gives good approximate solutions at the same time for all norms. The L_2 or L_3-norm are useful since they describe the performance of the whole schedule from the administrator point of view as well as they provide a fairness indication to the users. The hard point here is to derive theoretical analysis for such complicated tools.

Resource Augmentation. The classical resource augmentation models, i.e. speed and machine augmentation, are not sufficient to get good results when the execution of jobs cannot be frequently interrupted. However, based on a resource augmentation model recently introduced, where the algorithm may reject a small number of jobs, some members of our team have given the first interesting results in the non-preemptive direction. In general, resource augmentation can explain the intuitive good behavior of some greedy algorithms while, more interestingly, it can give ideas for new algorithms. For example, in the rejection context we could dedicate a small number of nodes for the usually problematic rejected jobs. Some initial experiments show that this can lead to a schedule for the remaining jobs that is very close to the optimal one.

4.2. Empirical Studies of Large Scale Platforms

Experiments or realistic simulations are required to take into account the impact of allocations and assess the real behavior of scheduling algorithms. While theoretical models still have their interest to lay the groundwork for algorithmic designs, the models are necessarily reflecting a purified view of the reality. As transferring our algorithm in a more practical setting is an important part of our creed, we need to ensure that the theoretical results found using simplified models can really be transposed to real situations. On the way to exascale computing, large scale systems become harder to study, to develop or to calibrate because of the costs in both time and energy of such processes. It is often impossible to convince managers to use a production cluster for several hours simply to test modifications in the RJMS. Moreover, as the existing RJMS production systems need to be highly reliable, each evolution requires several real scale test iterations. The consequence is that scheduling algorithms used in production systems are mostly outdated and not customized correctly. To circumvent this pitfall, we need to develop tools and methodologies for alternative empirical studies, from analysis of workload traces, to job models, simulation and emulation with reproducibility concerns.

4.2.1. *Workload Traces with Resource Consumption*

Workload traces are the base element to capture the behavior of complete systems composed of submitted jobs, running applications, and operating tools. These traces must be obtained on production platforms to provide relevant and representative data. To get a better understanding of the use of such systems, we need to look at both, how the jobs interact with the job management system, and how they use the allocated resources. We propose a general workload trace format that adds jobs resource consumption to the commonly used SWF [0] workload trace format. This requires to instrument the platforms, in particular to trace resource consumptions like CPU, data movements at memory, network and I/O levels, with an acceptable performance impact. In a previous work we studied and proposed a dedicated job monitoring tool whose impact on the system has been measured as lightweight (0.35% speed-down) with a 1 minute sampling rate. Other tools also explore job monitoring, like TACC Stats. A unique feature from our tool is its ability to monitor distinctly jobs sharing common nodes.

---

[0]Standard Workload Format: http://www.cs.huji.ac.il/labs/parallel/workload/swf.html

Collected workload traces with jobs resource consumption will be publicly released and serve to provide data for works presented in Section 4.1 . The trace analysis is expected to give valuable insights to define models encompassing complex behaviours like network topology sensitivity, network congestion and resource interferences.

We expect to join efforts with partners for collecting quality traces (ATOS/Bull, Ciment meso center, Joint Laboratory on Extreme Scale Computing) and will collaborate with the Inria team POLARIS for their analysis.

4.2.2. *Simulation*

Simulations of large scale systems are faster by multiple orders of magnitude than real experiments. Unfortunately, replacing experiments with simulations is not as easy as it may sound, as it brings a host of new problems to address in order to ensure that the simulations are closely approximating the execution of typical workloads on real production clusters. Most of these problems are actually not directly related to scheduling algorithms assessment, in the sense that the workload and platform models should be defined independently from the algorithm evaluations, in order to ensure a fair assessment of the algorithms' strengths and weaknesses. These research topics (namely platform modeling, job models and simulator calibration) are addressed in the other subsections.

We developed an open source platform simulator within DataMove (in conjunction with the OAR development team) to provide a widely distributable test bed for reproducible scheduling algorithm evaluation. Our simulator, named Batsim, allows to simulate the behavior of a computational platform executing a workload scheduled by any given scheduling algorithm. To obtain sound simulation results and to broaden the scope of the experiments that can be done thanks to Batsim, we did not chose to create a (necessarily limited) simulator from scratch, but instead to build on top of the SimGrid simulation framework.

To be open to as many batch schedulers as possible, Batsim decouples the platform simulation and the scheduling decisions in two clearly-separated software components communicating through a complete and documented protocol. The Batsim component is in charge of simulating the computational resources behaviour whereas the scheduler component is in charge of taking scheduling decisions. The scheduler component may be both a resource and a job management system. For jobs, scheduling decisions can be to execute a job, to delay its execution or simply to reject it. For resources, other decisions can be taken, for example to change the power state of a machine i.e. to change its speed (in order to lower its energy consumption) or to switch it on or off. This separation of concerns also enables interfacing with potentially any commercial RJMS, as long as the communication protocol with Batsim is implemented. A proof of concept is already available with the OAR RJMS.

Using this test bed opens new research perspectives. It allows to test a large range of platforms and workloads to better understand the real behavior of our algorithms in a production setting. In turn, this opens the possibility to tailor algorithms for a particular platform or application, and to precisely identify the possible shortcomings of the theoretical models used.

4.2.3. *Job and Platform Models*

The central purpose of the Batsim simulator is to simulate job behaviors on a given target platform under a given resource allocation policy. Depending on the workload, a significant number of jobs are parallel applications with communications and file system accesses. It is not conceivable to simulate individually all these operations for each job on large plaforms with their associated workload due to implied simulation complexity. The challenge is to define a coarse grain job model accurate enough to reproduce parallel application behavior according to the target platform characteristics. We will explore models similar to the BSP (Bulk Synchronous Program) approach that decomposes an application in local computation supersteps ended by global communications and a global synchronization. The model parameters will be established by means of trace analysis as discussed previously, but also by instrumenting some parallel applications to capture communication patterns. This instrumentation will have a significant impact on the concerned application performance, restricting its use to a few applications only. There are a lot of recurrent applications executed on HPC platform, this fact will help to reduce the required number of instrumentations and captures. To assign

each job a model, we are considering to adapt the concept of application signatures as proposed in. Platform models and their calibration are also required. Large parts of these models, like those related to network, are provided by Simgrid. Other parts as the filesystem and energy models are comparatively recent and will need to be enhanced or reworked to reflect the HPC platform evolutions. These models are then generally calibrated by running suitable benchmarks.

4.2.4. *Emulation and Reproducibility*

The use of coarse models in simulation implies to set aside some details. This simplification may hide system behaviors that could impact significantly and negatively the metrics we try to enhance. This issue is particularly relevant when large scale platforms are considered due to the impossibility to run tests at nominal scale on these real platforms. A common approach to circumvent this issue is the use of emulation techniques to reproduce, under certain conditions, the behavior of large platforms on smaller ones. Emulation represents a natural complement to simulation by allowing to execute directly large parts of the actual evaluated software and system, but at the price of larger compute times and a need for more resources. The emulation approach was chosen in to compare two job management systems from workload traces of the CURIE supercomputer (80000 cores). The challenge is to design methods and tools to emulate with sufficient accuracy the platform and the workload (data movement, I/O transfers, communication, applications interference). We will also intend to leverage emulation tools like Distem from the MADYNES team. It is also important to note that the Batsim simulator also uses emulation techniques to support the core scheduling module from actual RJMS. But the integration level is not the same when considering emulation for larger parts of the system (RJMS, compute node, network and filesystem).

Replaying traces implies to prepare and manage complex software stacks including the OS, the resource management system, the distributed filesystem and the applications as well as the tools required to conduct experiments. Preparing these stacks generate specific issues, one of the major one being the support for reproducibility. We propose to further develop the concept of reconstructability to improve experiment reproducibility by capturing the build process of the complete software stack. This approach ensures reproducibility over time better than other ways by keeping all data (original packages, build recipe and Kameleon engine) needed to build the software stack.

In this context, the Grid'5000 (see Sec. 6.4) experimentation infrastructure that gives users the control on the complete software stack is a crucial tool for our research goals. We will pursue our strong implication in this infrastructure.

4.3. Integration of High Performance Computing and Data Analytics

Data produced by large simulations are traditionally handled by an I/O layer that moves them from the compute cores to the file system. Analysis of these data are performed after reading them back from files, using some domain specific codes or some scientific visualisation libraries like VTK. But writing and then reading back these data generates a lot of data movements and puts under pressure the file system. To reduce these data movements, **the in situ analytics paradigm proposes to process the data as closely as possible to where and when the data are produced**. Some early solutions emerged either as extensions of visualisation tools or of I/O libraries like ADIOS. But significant progresses are still required to provide efficient and flexible high performance scientific data analysis tools. Integrating data analytics in the HPC context will have an impact on resource allocation strategies, analysis algorithms, data storage and access, as well as computer architectures and software infrastructures. But this paradigm shift imposed by the machine performance also sets the basis for a deep change on the way users work with numerical simulations. The traditional workflow needs to be reinvented to make HPC more user-centric, more interactive and turn HPC into a commodity tool for scientific discovery and engineering developments. In this context DataMove aims at investigating programming environments for in situ analytics with a specific focus on task scheduling in particular, to ensure an efficient sharing of resources with the simulation.

4.3.1. Programming Model and Software Architecture

In situ creates a tighter loop between the scientist and her/his simulation. As such, an in situ framework needs to be flexible to let the user define and deploy its own set of analysis. A manageable flexibility requires to favor simplicity and understandability, while still enabling an efficient use of parallel resources. Visualization libraries like VTK or Visit, as well as domain specific environments like VMD have initially been developed for traditional post-mortem data analysis. They have been extended to support in situ processing with some simple resource allocation strategies but the level of performance, flexibility and ease of use that is expected requires to rethink new environments. There is a need to develop a middleware and programming environment taking into account in its fundations this specific context of high performance scientific analytics.

Similar needs for new data processing architectures occurred for the emerging area of Big Data Analytics, mainly targeted to web data on cloud-based infrastructures. Google Map/Reduce and its successors like Spark or Stratosphere/Flink have been designed to match the specific context of efficient analytics for large volumes of data produced on the web, on social networks, or generated by business applications. These systems have mainly been developed for cloud infrastructures based on commodity architectures. They do not leverage the specifics of HPC infrastructures. Some preliminary adaptations have been proposed for handling scientific data in a HPC context. However, these approaches do not support in situ processing.

Following the initial development of FlowVR, our middleware for in situ processing, we will pursue our effort to develop a programming environment and software architecture for high performance scientific data analytics. Like FlowVR, the map/reduce tools, as well as the machine learning frameworks like TensorFlow, adopted a dataflow graph for expressing analytics pipe-lines. We are convinced that this dataflow approach is both easy to understand and yet expresses enough concurrency to enable efficient executions. The graph description can be compiled towards lower level representations, a mechanism that is intensively used by Stratosphere/Flink for instance. Existing in situ frameworks, including FlowVR, inherit from the HPC way of programming with a thiner software stack and a programming model close to the machine. Though this approach enables to program high performance applications, this is usually too low level to enable the scientist to write its analysis pipe-line in a short amount of time. The data model, i.e. the data semantics level accessible at the framework level for error check and optimizations, is also a fundamental aspect of such environments. The key/value store has been adopted by all map/reduce tools. Except in some situations, it cannot be adopted as such for scientific data. Results from numerical simulations are often more structured than web data, associated with acceleration data structures to be processed efficiently. We will investigate data models for scientific data building on existing approaches like Adios or DataSpaces.

4.3.2. Resource Sharing

To alleviate the I/O bottleneck, the in situ paradigm proposes to start processing data as soon as made available by the simulation, while still residing in the memory of the compute node. In situ processings include data compression, indexing, computation of various types of descriptors (1D, 2D, images, etc.). Per se, reducing data output to limit I/O related performance drops or keep the output data size manageable is not new. Scientists have relied on solutions as simple as decreasing the frequency of result savings. In situ processing proposes to move one step further, by providing a full fledged processing framework enabling scientists to more easily and thoroughly manage the available I/O budget.

The most direct way to perform in situ analytics is to inline computations directly in the simulation code. In this case, in situ processing is executed in sequence with the simulation that is suspended meanwhile. Though this approach is direct to implement and does not require complex framework environments, it does not enable to overlap analytics related computations and data movements with the simulation execution, preventing to efficiently use the available resources. Instead of relying on this simple time sharing approach, several works propose to rely on space sharing where one or several cores per node, called *helper cores*, are dedicated to analytics. The simulation responsibility is simply to handle a copy of the relevant data to the node-local in situ processes, both codes being executed concurrently. This approach often lead to significantly beter performance than in-simulation analytics.

For a better isolation of the simulation and in situ processes, one solution consists in offloading in situ tasks from the simulation nodes towards extra dedicated nodes, usually called *staging nodes*. These computations are said to be performed *in-transit*. But this approach may not always be beneficial compared to processing on simulation nodes due to the costs of moving the data from the simulation nodes to the staging nodes.

FlowVR enables to mix these different resources allocation strategies for the different stages of an analytics pile-line. Based on a component model, the scientist designs analytics workflows by first developing processing components that are next assembled in a dataflow graph through a Python script. At runtime the graph is instantiated according to the execution context, FlowVR taking care of deploying the application on the target architecture, and of coordinating the analytics workflows with the simulation execution.

But today the choice of the resource allocation strategy is mostly ad-hoc and defined by the programmer. We will investigate solutions that enable a cooperative use of the resource between the analytics and the simulation with minimal hints from the programmer. In situ processings inherit from the parallelization scale and data distribution adopted by the simulation, and must execute with minimal perturbations on the simulation execution (whose actual resource usage is difficult to know a priori). We need to develop adapted scheduling strategies that operate at compile and run time. Because analysis are often data intensive, such solutions must take into consideration data movements, a point that classical scheduling strategies designed first for compute intensive applications often overlook. We expect to develop new scheduling strategies relying on the methodologies developed in Sec. 4.1.5 . Simulations as well as analysis are iterative processes exposing a strong spatial and temporal coherency that we can take benefit of to anticipate their behavior and then take more relevant resources allocation strategies, possibly based on advanced learning algorithms or as developed in Section 4.1 .

In situ analytics represent a specific workload that needs to be scheduled very closely to the simulation, but not necessarily active during the full extent of the simulation execution and that may also require to access data from previous runs (stored in the file system or on specific burst-buffers). Several users may also need to run concurrent analytics pipe-lines on shared data. This departs significantly from the traditional batch scheduling model, motivating the need for a more elastic approach to resource provisioning. These issues will be conjointly addressed with research on batch scheduling policies (Sec. 4.1).

4.3.3. *Co-Design with Data Scientists*

Given the importance of users in this context, it is of primary importance that in situ tools be co-designed with advanced users, even if such multidisciplinary collaborations are challenging and require constant long term investments to learn and understand the specific practices and expectations of the other domain.

We will tightly collaborate with scientists of some application domains, like molecular dynamics or fluid simulation, to design, develop, deploy and assess in situ analytics scenarios, as already done with Marc Baaden, a computational biologist from LBT.

We recently extended our collaboration network. We started in 2015 a PhD co-advised with CEA DAM to investigate in situ analytics scenarios in the context of atomistic material simulations. CEA DAM is a French energy lab hosting one of the largest european supercomputer. They gather physicists, numerical scientists as well as high performance computer engineers, making it a very interesting partner for developing new scientific data analysis solutions. We also got a national grant (2015-2018) to compute in situ statistics for multi-parametric parallel studies with the research department of French power company EDF. In this context we collaborate with statisticians and fluid simulation experts to define in situ scenarios, revisit the statistic operators to be amenable to in situ processing, and define an adapted in situ framework.

DIANA Project-Team (section vide)

DIONYSOS Project-Team

4. Application Domains

4.1. Networking

Our global research effort concerns networking problems, both from the analysis point of view, and around network design issues. Specifically, this means the IP technology in general, with focus on specific types of networks seen at different levels: wireless systems, optical infrastructures, peer-to-peer architectures, Software Defined Networks, Content Delivery Networks, Content-Centric Networks, clouds.

A specific aspect of network applications and/or services based on video or voice content, is our PSQA technology, able to measure the Perceptual Quality automatically and in real time. PSQA provides a MOS value as close as it makes sense to the value obtained from subjective testing sessions. The technology has been tested in many environments, including one way communications as, for instance, in video streaming, and bi-directional communications as in IP telephony, UDP- or TCP-based systems, etc. It has already served in many collaborative projects as the measuring tool used.

4.2. Stochastic modeling

Many of the techniques developed at Dionysos are related to the analysis of complex systems in general, not only in telecommunications. For instance, our Monte Carlo methods for analyzing rare events have been used by different industrial partners, some of them in networking but recently also by companies building transportation systems. We develop methods in different areas: numerical analysis of stochastic models, bound computations in the same area, Discrete Event Simulation, or, as just mentioned, rare event analysis.

<p style="text-align:center;color:red;font-weight:bold;">DIVERSE Project-Team</p>

4. Application Domains

4.1. From Embedded Systems to Service Oriented Architectures

From small embedded systems such as home automation products or automotive systems to medium sized systems such as medical equipment, office equipment, household appliances, smart phones; up to large Service Oriented Architectures (SOA), building a new application from scratch is no longer possible. Such applications reside in (group of) machines that are expected to run continuously for years without unrecoverable errors. Special care has then to be taken to design and validate embedded software, making the appropriate trade-off between various extra-functional properties such as reliability, timeliness, safety and security but also development and production cost, including resource usage of processor, memory, bandwidth, power, etc.

Leveraging ongoing advances in hardware, embedded software is playing an evermore crucial role in our society, bound to increase even more when embedded systems get interconnected to deliver ubiquitous SOA. For this reason, embedded software has been growing in size and complexity at an exponential rate for the past 20 years, pleading for a component based approach to embedded software development. There is a real need for flexible solutions allowing to deal at the same time with a wide range of needs (product lines modeling and methodologies for managing them), while preserving quality and reducing the time to market (such as derivation and validation tools).

We believe that building flexible, reliable and efficient embedded software will be achieved by reducing the gap between executable programs, their models, and the platform on which they execute, and by developing new composition mechanisms as well as transformation techniques with a sound formal basis for mapping between the different levels.

Reliability is an essential requirement in a context where a huge number of softwares (and sometimes several versions of the same program) may coexist in a large system. On one hand, software should be able to evolve very fast, as new features or services are frequently added to existing ones, but on the other hand, the occurrence of a fault in a system can be very costly, and time consuming. While we think that formal methods may help solving this kind of problems, we develop approaches where they are kept "behind the scene" in a global process taking into account constraints and objectives coming from user requirements.

Software testing is another aspect of reliable development. Testing activities mostly consist in trying to exhibit cases where a system implementation does not conform to its specifications. Whatever the efforts spent for development, this phase is of real importance to raise the confidence level in the fact that a system behaves properly in a complex environment. We also put a particular emphasis on on-line approaches, in which test and observation are dynamically computed during execution.

DYOGENE Project-Team

4. Application Domains

4.1. Physical communication networks

Internet, wireless, mobile, cellular networks.

4.2. Abstract networks

Social interactions, human communities, economic networks.

4.3. Power grids

Energy networks.

<p style="text-align:center">**EVA Project-Team**</p>

4. Application Domains

4.1. Industrial Process Automation

Wireless networks have become ubiquitous and are an integral part of our daily lives. These networks are present in many application domains; the most important are detailed in this section.

Networks in industrial process automation typically perform **monitoring and control** tasks. Wired industrial communication networks, such as HART [0], have been around for decades and, being wired, are highly reliable. Network administrators tempted to "go wireless" expect the same reliability. Reliable process automation networks – especially when used for control – often impose stringent latency requirements. Deterministic wireless networks can be used in critical systems such as control loops, however, the unreliable nature of the wireless medium, coupled with their large scale and "ad-hoc" nature raise some of the most important challenges for low-power wireless research over the next 5-10 years.

Through the involvement of team members in standardization activities, the protocols and techniques will be proposed for the standardization process with a view to becoming the *de-facto* standard for wireless industrial process automation. Besides producing top level research publications and standardization activities, EVA intends this activity to foster further collaborations with industrial partners.

4.2. Environmental Monitoring

Today, outdoor WSNs are used to monitor vast rural or semi-rural areas and may be used to detect fires. Another example is detecting fires in outdoor fuel depots, where the delivery of alarm messages to a monitoring station in an upper-bounded time is of prime importance. Other applications consist in monitoring the snow melting process in mountains, tracking the quality of water in cities, registering the height of water in pipes to foresee flooding, etc. These applications lead to a vast number of technical issues: deployment strategies to ensure suitable coverage and good network connectivity, energy efficiency, reliability and latency, etc.

We work on such applications in an associate team "REALMS" comprising members from EVA, the university of Berkeley and the university of Michigan.

4.3. The Internet of Things

The general agreement is that the Internet of Things (IoT) is composed of small, often battery-powered objects which measure and interact with the physical world, and encompasses smart home applications, wearables, smart city and smart plant applications.

It is absolutely essential to (1) clearly understand the limits and capabilities of the IoT, and (2) develop technologies which enable user expectation to be met.

The EVA team is dedicated to understanding and contributing to the IoT. In particular, the team will maintain a good understanding of the different technologies at play (Bluetooth, IEEE 802.15.4, WiFi, cellular), and their trade-offs. Through scientific publications and other contributions, EVA will help establishing which technology best fits which application.

---

[0]Highway Addressable Remote Transducer

4.4. Military, Energy and Aerospace

Through the HIPERCOM project, EVA has developed cutting-edge expertise in using wireless networks for military, energy and aerospace applications. Wireless networks are a key enabling technology in the application domains, as they allow physical processes to be instrumented (e.g. the structural health of an airplane) at a granularity not achievable by its wired counterpart. Using wireless technology in these domains does however raise many technical challenges, including end-to-end latency, energy-efficiency, reliability and Quality of Service (QoS). Mobility is often an additional constraint in energy and military applications. Achieving scalability is of paramount importance for tactical military networks, and, albeit to a lesser degree, for power plants. EVA will work in this domain.

Smart cities share the constraint of mobility (both pedestrian and vehicular) with tactical military networks. Vehicular Ad-hoc NETworks (VANETs) will play an important role in the development of smarter cities.

The coexistence of different networks operating in the same radio spectrum can cause interference that should be avoided. Cognitive radio provides secondary users with the frequency channels that are temporarily unused (or unassigned) by primary users. Such opportunistic behavior can also be applied to urban wireless sensor networks. Smart cities raise the problem of transmitting, gathering, processing and storing big data. Another issue is to provide the right information at the place where it is most needed.

4.5. Emergency Applications

In an "emergency" application, heterogeneous nodes of a wireless network cooperate to recover from a disruptive event in a timely fashion, thereby possibly saving human lives. These wireless networks can be rapidly deployed and are useful to assess damage and take initial decisions. Their primary goal is to maintain connectivity with the humans or mobile robots (possibly in a hostile environment) in charge of network deployment. The deployment should ensure the coverage of particular points or areas of interest. The wireless network has to cope with pedestrian mobility and robot/vehicle mobility. The environment, initially unknown, is progressively discovered and may contain numerous obstacles that should be avoided. The nodes of the wireless network are usually battery-powered. Since they are placed by a robot or a human, their weight is very limited. The protocols supported by these nodes should be energy-efficient to maximize network lifetime. In such a challenging environment, sensor nodes should be replaced before their batteries are depleted. It is therefore important to be able to accurately determine the battery lifetime of these nodes, enabling predictive maintenance.

4.6. Types of Wireless Networks

The EVA team will distinguish between opportunistic communication (which takes advantage of a favorable state) and = collaborative communication (several entities collaborate to reach a common objective). Furthermore, determinism can be required to schedule medium access and node activity, and to predict energy consumption.

In the EVA project, we will propose **self-adaptive wireless networks** whose evolution is based on:

- optimization to minimize a single or multiple objective functions under some constraints (e.g. interference, or energy consumption in the routing process).
- machine learning to be able to predict a future state based on past states (e.g. link quality in a wireless sensor network) and to identify tendencies.

The types of wireless networks encountered in the application domains can be classified in the following categories.

4.6.1. Wireless Sensor and Mesh Networks

Standardization activities at the IETF have defined an "upper stack" allowing low-power mesh networks to be seamlessly integrated in the Internet (6LoWPAN), form multi-hop topologies (RPL), and interact with other devices like regular web servers (CoAP).

Major research challenges in sensor networks are mostly related to (predictable) power conservation and efficient multi-hop routing. Applications such as monitoring of mobile targets, and the generalization of smart phone devices and wearables, have introduced the need for WSN communication protocols to cope with node mobility and intermittent connectivity.

Extending WSN technology to new application spaces (e.g. security, sports, hostile environments) could also assist communication by seamless exchanges of information between individuals, between individuals and machines, or between machines, leading to the Internet of Things.

4.6.2. *Deterministic Low-Power Networks*

Wired sensor networks have been used for decades to automate production processes in industrial applications, through standards such as HART. Because of the unreliable nature of the wireless medium, a wireless version of such industrial networks was long considered infeasible.

In 2012, the publication of the IEEE 802.15.4e standard triggered a revolutionary trend in low-power mesh networking: merging the performance of industrial networks, with the ease-of-integration of IP-enabled networks. This integration process is spearheaded by the IETF 6TiSCH working group, created in 2013. A 6TiSCH network implements the IEEE 802.15.4e TSCH protocol, as well as IETF standards such as 6LoWPAN, RPL and CoAP. A 6TiSCH network is synchronized, and a communication schedule orchestrates all communication in the network. Deployments of pre-6TiSCH networks have shown that they can achieve over 99.999% end-to-end reliability, and a decade of battery lifetime.

The communication schedule of a 6TiSCH network can be built and maintained using a centralized, distributed, or hybrid scheduling approach. While the mechanisms for managing that schedule are being standardized by the IETF, which scheduling approach to use, and the associated limits in terms of reliability, throughput and power consumption remains entirely open research questions. Contributing to answering these questions is an important research direction for the EVA team.

4.6.3. *MANETs and VANETs*

In contrast to routing, other domains in MANETs such as medium access, multi-carrier transmission, quality of service, and quality of experience have received less attention. The establishment of research contracts for EVA in the field of MANETs is expected to remain substantial. MANETs will remain a key application domain for EVA with users such as the military, firefighters, emergency services and NGOs.

Vehicular Ad hoc Networks (VANETs) are arguably one of the most promising applications for MANETs. These networks primarily aim at improving road safety. Radio spectrum has been ring-fenced for VANETs worldwide, especially for safety applications. International standardization bodies are working on building efficient standards to govern vehicle-to-vehicle or vehicle-to-infrastructure communication.

4.6.4. *Cellular and Device-to-Device Networks*

We propose to initially focus this activity on spectrum sensing. For efficient spectrum sensing, the first step is to discover the links (subcarriers) on which nodes may initiate communications.= In Device-to-Device (D2D) networks, one difficulty is scalability.

For link sensing, we will study and design new random access schemes for D2D networks, starting from active signaling. This will assume the availability of a control channel devoted to D2D neighbor discovery. It is therefore naturally coupled with cognitive radio algorithms (allocating such resources): coordination of link discovery through eNode-B information exchanges can yield further spectrum usage optimization.

<p style="text-align:center; color:red;">**FOCUS Project-Team**</p>

4. Application Domains

4.1. Ubiquitous Systems

The main application domain for Focus are ubiquitous systems, broadly systems whose distinctive features are: mobility, high dynamicity, heterogeneity, variable availability (the availability of services offered by the constituent parts of a system may fluctuate, and similarly the guarantees offered by single components may not be the same all the time), open-endedness, complexity (the systems are made of a large number of components, with sophisticated architectural structures). In Focus we are particularly interested in the following aspects.

- *Linguistic primitives* for programming dialogues among components.
- *Contracts* expressing the functionalities offered by components.
- *Adaptability and evolvability* of the behaviour of components.
- *Verification* of properties of component systems.
- Bounds on component *resource consumption* (e.g., time and space consumed).

4.2. Service Oriented Computing and Cloud Computing

Today the component-based methodology often refers to Service Oriented Computing. This is a specialized form of component-based approach. According to W3C, a service-oriented architecture is "a set of components which can be invoked, and whose interface descriptions can be published and discovered". In the early days of Service Oriented Computing, the term services was strictly related to that of Web Services. Nowadays, it has a much broader meaning as exemplified by the XaaS (everything as a service) paradigm: for example, based on modern virtualization technologies, cloud computing offers the possibility to build sophisticated service systems on virtualized infrastructures accessible from everywhere and from any kind of computing device. Such infrastructures are usually examples of sophisticated service oriented architectures that, differently from traditional service systems, should also be capable to elastically adapt on demand to the user requests.

<p style="text-align: center; color: red;">**FUN Project-Team**</p>

4. Application Domains

4.1. Application Domains

The set of applications enabled through FUN and IoT is very large and can apply in every application area. We can thus not be exhaustive but among the most spread applications, we can name every area, event or animal monitoring, understanding and protection. To illustrate this, we may refer to the use cases addressed by our PREDNET project which goals is to equip rhinoceros with smart communicating devices to fight against poaching.

Other field of application is exploration of hostile and/or unknown environment by a fleet of self-organizing robots that cooperate with RFID and sensors to ensure a continue monitoring afterwards.

Also, IoT and FUN ca play a key role in logistics and traceability by relying on the use of sensors or RFID technologies as implemented in our TRACAVERRE project or our collaboration with the start up TRAXENS.

Finally, IoT and FUN leverage a lot of applications in Smart City concept , ranging from parking aid to a better energy consumption going through air quality monitoring, traffic fluidizing etc. (See our CityLab Inria and VITAL projects).

GANG Project-Team

4. Application Domains

4.1. Large scale networks

Application domains include evaluating Internet performances, the design of new peer-to-peer applications, enabling large scale networks, and developping tools for transportation networks.

HIEPACS Project-Team

4. Application Domains

4.1. Material physics

Participants: Pierre Blanchard, Olivier Coulaud, Arnaud Durocher.

Due to the increase of available computer power, new applications in nano science and physics appear such as study of properties of new materials (photovoltaic materials, bio- and environmental sensors, ...), failure in materials, nano-indentation. Chemists, physicists now commonly perform simulations in these fields. These computations simulate systems up to billion of atoms in materials, for large time scales up to several nanoseconds. The larger the simulation, the smaller the computational cost of the potential driving the phenomena, resulting in low precision results. So, if we need to increase the precision, there are two ways to decrease the computational cost. In the first approach, we improve algorithms and their parallelization and in the second way, we will consider a multiscale approach.

A domain of interest is the material aging for the nuclear industry. The materials are exposed to complex conditions due to the combination of thermo-mechanical loading, the effects of irradiation and the harsh operating environment. This operating regime makes experimentation extremely difficult and we must rely on multi-physics and multi-scale modeling for our understanding of how these materials behave in service. This fundamental understanding helps not only to ensure the longevity of existing nuclear reactors, but also to guide the development of new materials for 4th generation reactor programs and dedicated fusion reactors. For the study of crystalline materials, an important tool is dislocation dynamics (DD) modeling. This multiscale simulation method predicts the plastic response of a material from the underlying physics of dislocation motion. DD serves as a crucial link between the scale of molecular dynamics and macroscopic methods based on finite elements; it can be used to accurately describe the interactions of a small handful of dislocations, or equally well to investigate the global behavior of a massive collection of interacting defects.

To explore i.e. to simulate these new areas, we need to develop and/or to improve significantly models, schemes and solvers used in the classical codes. In the project, we want to accelerate algorithms arising in those fields. We will focus on the following topics (in particular in the currently under definition OPTIDIS project in collaboration with CEA Saclay, CEA Ile-de-france and SIMaP Laboratory in Grenoble) in connection with research described at Sections 3.4 and 3.5 .

- The interaction between dislocations is long ranged ($O(1/r)$) and anisotropic, leading to severe computational challenges for large-scale simulations. In dislocation codes, the computation of interaction forces between dislocations is still the most CPU time consuming and has to be improved to obtain faster and more accurate simulations.

- In such simulations, the number of dislocations grows while the phenomenon occurs and these dislocations are not uniformly distributed in the domain. This means that strategies to dynamically construct a good load balancing are crucial to acheive high performance.

- From a physical and a simulation point of view, it will be interesting to couple a molecular dynamics model (atomistic model) with a dislocation one (mesoscale model). In such three-dimensional coupling, the main difficulties are firstly to find and characterize a dislocation in the atomistic region, secondly to understand how we can transmit with consistency the information between the two micro and meso scales.

4.2. Co-design for scalable numerical algorithms in scientific applications

Participants: Nicolas Bouzat, Mathieu Faverge, Guillaume Latu, Michel Mehrenberger, Pierre Ramet, Jean Roman.

4.2.1. High performance simulation for ITER tokamak

Scientific simulation for ITER tokamak modeling provides a natural bridge between theory and experimentation and is also an essential tool for understanding and predicting plasma behavior. Recent progresses in numerical simulation of fine-scale turbulence and in large-scale dynamics of magnetically confined plasma have been enabled by access to petascale supercomputers. These progresses would have been unreachable without new computational methods and adapted reduced models. In particular, the plasma science community has developed codes for which computer runtime scales quite well with the number of processors up to thousands cores. The research activities of HIEPACS concerning the international ITER challenge were involved in the Inria Project Lab C2S@EXA in collaboration with CEA-IRFM and are related to two complementary studies: a first one concerning the turbulence of plasma particles inside a tokamak (in the context of GYSELA code) and a second one concerning the MHD instability edge localized modes (in the context of JOREK code).

Currently, GYSELA is parallelized in an hybrid MPI+OpenMP way and can exploit the power of the current greatest supercomputers. To simulate faithfully the plasma physic, GYSELA handles a huge amount of data and today, the memory consumption is a bottleneck on very large simulations. In this context, mastering the memory consumption of the code becomes critical to consolidate its scalability and to enable the implementation of new numerical and physical features to fully benefit from the extreme scale architectures.

Other numerical simulation tools designed for the ITER challenge aim at making a significant progress in understanding active control methods of plasma edge MHD instability Edge Localized Modes (ELMs) which represent a particular danger with respect to heat and particle loads for Plasma Facing Components (PFC) in the tokamak. The goal is to improve the understanding of the related physics and to propose possible new strategies to improve effectiveness of ELM control techniques. The simulation tool used (JOREK code) is related to non linear MHD modeling and is based on a fully implicit time evolution scheme that leads to 3D large very badly conditioned sparse linear systems to be solved at every time step. In this context, the use of PaStiX library to solve efficiently these large sparse problems by a direct method is a challenging issue.

INDES Project-Team (section vide)

INFINE Project-Team (section vide)

KERDATA Project-Team (section vide)

MADYNES Team

4. Application Domains

4.1. Mobile and constrained networks

The results coming out from MADYNES can be applied to any dynamic infrastructure that contributes to the delivery of value added services. While this is a potentially huge application domain, we mainly focus at the network level on mobile devices and Internet of Things. We are investigating the provisioning, monitoring, configuration and performance management issues.

4.2. Dynamic services infrastructures

At the service level, dynamics is also increasing very fast. We apply the results of our work on autonomous management on infrastructures which support dynamic composition and for which self-instrumentation and management automation is required. The target service environments are:

- sensor networks,
- cyber-physical systems,
- information centric networks,
- distributed cloud environnements,
- smart environments.

<p style="text-align:center; color:red;">**MIMOVE Team**</p>

4. Application Domains

4.1. Mobile urban systems for smarter cities

With the massive scale adoption of mobile devices and further expected significant growth in relation with the Internet of Things, mobile computing is impacting most – if not all – the ICT application domains. However, given the importance of conducting empirical studies to assess and nurture our research, we focus on one application area that is the one of "*smart cities*". The smart city vision anticipates that the whole urban space, including buildings, power lines, gas lines, roadways, transport networks, and cell phones, can all be wired together and monitored. Detailed information about the functioning of the city then becomes available to both city dwellers and businesses, thus enabling better understanding and consequently management of the city's infrastructure and resources. This raises the prospect that cities will become more sustainable environments, ultimately enhancing the citizens' well being. There is the further promise of enabling radically new ways of living in, regulating, operating and managing cities, through the increasing active involvement of citizens by ways of crowd-sourcing/sensing and social networking.

Still, the vision of what smart cities should be about is evolving at a fast pace in close concert with the latest technology trends. It is notably worth highlighting how mobile and social network use has reignited citizen engagement, thereby opening new perspectives for smart cities beyond data analytics that have been initially one of the core foci for smart cities technologies. Similarly, open data programs foster the engagement of citizens in the city operation and overall contribute to make our cities more sustainable. The unprecedented democratization of urban data fueled by open data channels, social networks and crowd sourcing enables not only the monitoring of the activities of the city but also the assessment of their nuisances based on their impact on the citizens, thereby prompting social and political actions. However, the comprehensive integration of urban data sources for the sake of sustainability remains largely unexplored. This is an application domain that we intend to focus on, further leveraging our research on emergent mobile distributed systems, large-scale mobile sensing & actuation, and mobile social crowd-sensing.

In a first step, we concentrate on the following specialized applications, which we investigate in close collaboration with other researchers, in particular as part of the dedicated Inria Project Lab *CityLab@Inria*:

- **Democratization of urban data for healthy cities.** The objective here is to integrate the various urban data sources, especially by way of crowd-Xing, to better understand city nuisances from raw pollution sensing (e.g., sensing noise) to the sensing of its impact on citizens (e.g., how people react to urban noise and how this affects their health).

- **Socially-aware urban mobility.** Mobility within mega-cities is known as one of the major challenges to face urgently due to the fact that today's mobility patterns do not scale and to the negative effect on the environment and health. It is our belief that mobile social and physical sensing may significantly help in promoting the use of public transport, which we have started to investigate through empirical study based on the development and release of dedicated apps.

- **Social applications.** Mobile applications are being considered by sociologists as a major vehicle to actively involve citizens and thereby prompt them to become activists. This is especially studied with the Social Apps Lab at UC Berkeley. Our objective is to study such a vehicle from the ICT perspective and in particular elicit relevant middleware solutions to ease the development and development of such "*civic apps*".

Acknowledging the need for collaborative research in the application domain of smart cities, MiMove is heavily involved and actually leading CityLab@Inria [0]. The Inria Project Lab CityLab is focused on the study of ICT solutions promoting social sustainability in smart cities, and involves the following Inria project-teams in addition to MiMove: CLIME, DICE, FUN, MYRIADS, SMIS, URBANET and WILLOW. CityLab further involves strong collaboration with California universities affiliated with CITRIS (Center for Information Technology Research in the Interest of Society) and especially UC Berkeley, in relation with the *Inria@SiliconValley* program. We note that Valérie Issarny acts as scientific manager of Inria@SiliconValley and is currently visiting scholar at CITRIS at UC Berkeley. In this context, MiMove researchers are working closely with colleagues of UC Berkeley, including researchers from various disciplines interested in smart cities (most notably sociologists).

4.2. Home Network Diagnosis

With the availability of cheap broadband connectivity, Internet access from the home has become a ubiquity. Modern households host a multitude of networked devices, ranging from personal devices such as laptops and smartphones to printers and media centers. These devices connect among themselves and to the Internet via a local-area network —a *home network*– that has become an important part of the "Interne experience". In fact, ample anecdotal evidence suggests that the home network can cause a wide array of connectivity impediments, but their nature, prevalence, and significance remain largely unstudied.

Our long-term goal is to assist users with concrete indicators of the causes of potential problems and—ideally—ways to fix them. We intend to develop a set of easy-to-use home network diagnosis tools that can reliably identify performance and functionality shortcomings rooted in the home. The development of home network diagnosis tools brings a number of challenges. First, home networks are heterogeneous. The set of devices, configurations, and applications in home networks vary significantly from one home to another. We must develop sophisticated techniques that can learn and adapt to any home network as well as to the level of expertise of the user. Second, there are numerous ways in which applications can fail or experience poor performance in home networks. Often there are a number of explanations for a given symptom. We must devise techniques that can identify the most likely cause(s) for a given problem from a set of possible causes. Third, even if we can identify the cause of the problem, we must then be able to identify a solution. It is important that the output of the diagnosis tools we build is "actionable". Users should understand the output and know what to do.

We are conceiving methods for two application scenarios: (i) when the end user in the home deploys our diagnostic tools either on the home gateway (the gateway often combines a DSL/cable modem and an access point; it connects the home network to the ISP) or on devices connected to the home network and (ii) when ISPs collect measurements from homes of subscribers and then correlate these measurements to help identify problems.

Assisting end users. We are developing algorithms to determine whether network performance problems lie inside or outside the home network. Given that the home gateway connects the home with the rest of the Internet, we are designing an algorithm (called *HoA*) that analyzes traffic that traverses the gateway to distinguish access link and home network bottlenecks. A measurement vantage point on the gateway is key for determining if the performance bottleneck lies within the home network or the access ISP, but we also need to deploy diagnosis tools in end-devices. First, some users may not want (or not know how) to deploy a new home gateway in their homes. Second, some problems will be hard to diagnose with only the vantage point of the gateway (for example, when a device cannot send traffic or when the wireless is poor in certain locations of a home). We can obtain more complete visibility by leveraging *multiple* measurement nodes around the home, potentially including the home gateway, all participating jointly in the measurement task. We have an ongoing project to realize a home network analyzer as a web-based measurement application built on top of our team's recently developed browser-based measurement platform, *Fathom*. To integrate the home gateway in the analyzer, we plan to engage the BISmark Project. BISmark already provides a web server as well as

---

[0] http://citylab.inria.fr

extensive configurability, allowing us to experiment freely with both passive as well as active measurements. We must develop a home network analyzer that can first discover the set of devices connected to the home network that can collaborate on the diagnosis task. We will then develop tomography algorithms to infer where performance problems lie given measurements taken from the set of available vantage points.

Assisting Internet Service Providers (ISPs). Our discussions with several large access ISPs reveal that service calls are costly, ranging from $9–25 per call, and as many as 75% of service calls from customers are usually caused by problems that have nothing to do with the ISP. Therefore, ISPs are eager to deploy techniques to assist in home network diagnosis. In many countries ISPs control the home gateway and set-top-boxes in the home. We plan to develop more efficient mechanisms for home users to report trouble to their home ISP and consequently reduce the cost of service calls. This project is in collaboration with Technicolor and Portugal Telecom. Technicolor is a large manufacturer of home gateways and set-top-boxes. Portugal Telecom is the largest broadband access provider in Portugal. Technicolor already collects data from 200 homes in Portugal. We are working with the data collected in this deployment together with controlled experiments to develop methods to diagnose problems in the home wireless.

4.3. Quality of Experience

An increasing number of residential users consume online services (e.g., VoD, Web browsing, or Skype) in their everyday activities (e.g., for education or entertainment purposes), using a variety of devices (e.g., tablets, smartphones, laptops). A high Quality of Service (QoS) is essential for sustaining the revenue of service providers, carriers, and device manufactures. Yet, the perceived Quality of Experience (QoE) of users is far from perfect e.g., videos that get stalled or that take a long time to load. Dissatisfied users may change Internet Service Providers (ISPs) or the online services. Hence, the incentives for measuring and improving QoE in home networks are high while mapping network and application QoS to QoE is a challenging problem. In this work we have focused in measuring several network Quality-of-Service (QoS) metrics, such as latency and bandwidth, both in residential Wi-Fi as well as broadband networks, homes are using for connecting to the Internet.

The WiFi Context. Residential Wi-Fi performance, however, is highly variable. Competing Wi-Fi networks can cause contention and interference while poor channel conditions between the station and the access point (AP) can cause frame losses and low bandwidth. In some cases, the home Wi-Fi network can bottleneck Internet access. While problems in the Wi-Fi network may affect several network QoS metrics, users will typically only notice a problem when poor Wi-Fi affects the QoE of Internet applications. For example, a Wi-Fi network with low bandwidth may go unnoticed unless the time to load Web pages increases significantly. A user observing degraded QoE due to Wi-Fi problems may mistakenly assume there is a problem with the Internet Service Provider (ISP) network. Our discussions with residential ISPs confirm that often customers call to complain about problems in the home Wi-Fi and not the ISP network.

Prior work has focused on QoS metrics for some applications (e.g., on-line video, Web browsing, or Skype) with no attempt to identify when Wi-Fi quality affects QoE. We are particularly interested in assisting ISPs to predict when home Wi-Fi quality degrades QoE. ISPs can use this system to detect customers experiencing poor QoE to proactively trigger Wi-Fi troubleshooting. ISPs often control the home AP, so we leverage Wi-Fi metrics that are available on commercial APs. Detecting when Wi-Fi quality degrades QoE using these metrics is challenging. First, we have no information about the applications customers are running at any given time. ISPs avoid capturing per-packet traffic traces from customers, because of privacy considerations and the overload of per-packet capture. Thus, we must estimate the effect of Wi-Fi quality on QoE of popular applications, which most customers are likely to run. In this context, we study Web as a proof of concept, as a large fraction of home traffic corresponds to Web. Second, application QoE may be degraded by factors other than the Wi-Fi quality (e.g., poor Internet performance or an overloaded server). Although a general system to explain any QoE degradation would be extremely helpful, our monitoring at the AP prevents us from having the end-to-end view necessary for such general task. Instead, we focus on identifying when Wi-Fi quality degrades QoE. Finally, Wi-Fi metrics available in APs are coarse aggregates such as the average PHY rate or the fraction of busy times. It is open how to effectively map these coarse metrics into QoE.

Predicting QoE. Clearly, different actors in the online service chain (e.g., video streaming services, ISPs) have different incentives and means to measure and affect the user QoE. Uncovering statistically equivalent subsets of QoS metrics across and within levels provides actionable knowledge for building QoE predictors. To achieve this goal, we leverage recent advances on feature selection algorithms to exploit available experimental evidence of the joint probability distributions of QoE/QoS metrics. This type of statistical reasoning will enable us to determine local causal relationships between a target QoE variable, seen as effect, and multiple QoS metrics across or within levels, seen as causes. Such data-driven analysis is justified by the multiplicity of dependencies that exist between network or application QoS metrics as different adaptation mechanisms (e.g., TCP congestion avoidance, HTTP bitrate adaptation) are activated at each level in real life. Building optimal predictors based on (eventually several) probabilistically minimal subsets of features opens the way for a principled comparison of the predictors.

4.4. Crowd-sourced Information Filtering and Summarization

With the explosion of the People-centric Web, there is a proliferation of crowd-sourced content either under the form of qualitative reviews (mainly textual) and quantitative ratings (as 5 star ratings) regarding diverse products or services or under the form of various "real-time" feedback events (e.g., re-tweets, replies, likes, clicks, etc.) on published web content (ranging from traditional news, TV series, and movies to specialized blogs and posts shared over social networks). Such content captures the wisdom of the crowd and is valuable information source for building collaborative filtering systems and text summarization tools coping with information overload. For example, they can assist users to pick the most interesting web pages (e.g. Delicious) or to choose which movie to watch next (e.g. Netflix).

Implicit Feedback in Communities of a Place. We are initially interested in addressing one of the main limitation of collaborative filtering systems namely, the strong user engagement required to provide the necessary input (e.g., regarding their friends, tags or sites of preference) which is usual platform specific (i.e., for a particular social network, tagging, or bookmark system). The lack of user engagement translates into cold start and data sparsity. To cope with this limitation, we are developing a system called WeBrowse that passively observes network traffic to extract user clicks (i.e., the URLs users visit) for group of people who live, study, or work in the same place. Examples of such communities of a place are: (i) the students of a campus, (ii) the people living in a neighbourhood or (iii) researchers working in the same site. WeBrowse then promotes the hottest and most popular content to the community members sharing common interests.

Personalized Review Summarization. Finally, we are interested in helping people to take informed decisions regarding their shopping or entertainment activities. The automated summarization of a review corpus (for example, movie reviews from Rotten Tomatoes or IMDB; or restaurant reviews from Yelp) aims to assist people to form an opinion regarding a product/service of interest, by producing a coherent summary that is helpful and can be easily assimilated by humans. We are working on review summarisation methods that combine both objective (i.e., related to the review corpus) and subjective (i.e., related to the end-user interests) interestingness criteria of the produced reviews. In this respect we are exploiting domain models (e.g., Oscar's merit categories for movies) to elicit user preferences and mine the aspects of products/services actually commented in the textual sentences of reviews. For example, different summaries should be produced when a user is more interested in the actors' performance rather than the movie story. We are particularly interested in extracting automatically the signatures of aspects (based on a set of seed terms) and rank review sentences on their importance and relevance w.r.t. the aspects they comment. Last but not least we are optimizing the automatically constructed summary w.r.t. to a number of criteria such as the number of the length of included sentences from the original reviews, the polarity of sentiments in the described aspects, etc.

<p style="text-align:center;color:red;font-weight:bold;">MYRIADS Project-Team</p>

4. Application Domains

4.1. Main Application Domains

The Myriads team investigates the design and implementation of system services. Thus its research activities address a broad range of application domains. We validate our research results with selected use cases in the following application domains:

- Smart city services,
- Smart grids,
- Energy and sustainable development,
- Home IoT applications,
- Bio-informatics applications,
- Data science applications,
- Computational science applications,
- Numerical simulations.

<div align="center">**NEO Project-Team**</div>

4. Application Domains

4.1. Network Science

Network Science is a multidisciplinary body of knowledge, principally concerned with the emergence of global properties in a network of individual agents, from the "local" properties of this network, namely, the way agents interact with each other. The central model of "networks" is the graph (of Graph Theory/Operations Research), with nodes representing the different entities managing information and taking decisions, and the links representing the fact that entities interact, or not. Links are usually equipped with a "weight" that measures the intensity of interaction. Adding evolution rules to this quite elementary representation leads to dynamic network models, the properties of which Network Science tries to analyze.

A classical example of properties sought in networks is the famous "six degrees of separation" (or "small world") property: how and why does it happen so frequently? Another ubiquitous property of real-life networks is the Zipf or "scale-free" distribution for degrees. Some of these properties, when properly exploited, lead to successful business opportunities: just consider the PageRank algorithm of Google, which miraculously connects the relevance of some Web information with the relevance of the other information that points to it.

4.2. Network Engineering

In its primary acceptation, Network Science involves little or no engineering: phenomena are assumed to be "natural" and emerge without intervention. However, the idea comes fast to intervene in order to modify the outcome of the phenomenon. This is where NEO is positioned. Beyond the mostly descriptive approach of Network Science, we aim at using the techniques of Operations Research so as to engineer complex networks.

To quote just two examples: controlling the spread of diseases through a "network" of people is of primarily interest for mankind. Similarly, controlling the spread of information or reputation through a social network is of great interest in the Internet. Precisely: given the impact of web visibility on business income, it is tempting (and quite common) to manipulate the graph of the web by adding links so as to drive the PageRank algorithm to a desired outcome.

Another interesting example is the engineering of community structures. Recently, thousands of papers have been written on the topic of community *detection* problem. In most of the works, the researchers propose methods, most of the time, heuristics, for detecting communities or dense subgraphs inside a large network. Much less effort has been put in the understanding of community formation process and even much less effort has been dedicated to the question of how one can influence the process of community formation, e.g. in order to increase overlap among communities and reverse the fragmentation of the society.

Our ambition for the medium term is to reach an understanding of the behavior of complex networks that will make us capable of influencing or producing a certain property in said network. For this purpose, we will develop families of models to capture the essential structure, dynamics, and uncertainty of complex networks. The "solution" of these models will provide the correspondence between metrics of interest and model parameters, thus opening the way to the synthesis of effective control techniques.

In the process of tackling real, very large size networks, we increasingly deal with large graph data analysis and the development of decision techniques with low algorithmic complexity, apt at providing answers from large datasets in reasonable time.

<p style="text-align:center;">**PHOENIX Project-Team**</p>

4. Application Domains

4.1. Internet of Things

Participants: Charles Consel, Nic Volanschi.

The Internet of Things (IoT) has become a reality with the emergence of Smart Cities, populated with large amounts of smart objects which are used to deliver a range of citizen services (e.g., security, well being, etc.) The IoT paradigm relies on the pervasive presence of smart objects or "things", which raises a number of new challenges in the software engineering domain.

We introduce a design-driven development approach that is dedicated to the domain of orchestration of masses of sensors. The developer declares what an application does using a domain-specific language (DSL), named DiaSwarm. Our compiler processes domain-specific declarations to generate a customized programming framework that guides and supports the programming phase.

DiaSwarm addresses the main phases of an application orchestrating masses of sensors.

Service discovery Standard service discovery at the individual object level does not address the needs of applications orchestrating large numbers of smart objects. Instead, a high-level approach which provides constructs to specifying subsets of interest is needed. Our approach allows developers to introduce application-specific concepts (e.g., regrouping parking spaces into lots or districts) at the design time and then these can be used to express discovery operations. Following our design-driven development approach, these concepts are used to generate code to support and guide the programming phase.

Data gathering Applications need to acquire data from a large number of objects through a variety of delivery models. For instance, air pollution sensors across a city may only push data to the relevant applications when pollution levels exceed tolerated levels. Tracking sensors, however, might determine the location of vehicles and send the acquired measurements to applications periodically (e.g., 10 min. intervals). Data delivery models need to be introduced at design time since they have a direct impact on the application's program structure. In doing so, the delivery models used by an application can be checked against sensor features early in the development process.

Data processing Data that is generated from hundreds of thousands of objects and accumulated over a period of time calls for efficient processing strategies to ensure the required performance is attained. Our approach allows for an efficient implementation of the data processing stage by providing the developer with a framework based on the MapReduce [34] programming model which is intended for the processing of large data sets.

4.2. Assistive computing in the home

Participants: Quentin Barlas, Antoine Riche, Adrien Carteron, Rafik Belloum, Audrey Landuran, Julien Durand, Charlotte Froger, Stéphanie Giraud, Pauline Fontagne, Flora Gallet, Amandine Desrozier, Ludovic Fornasari, Léo Mendiboure, Hélène Sauzéon, Charles Consel, Bernard N'Kaoua, Bernard Serpette, Nic Volanschi.

In this avenue of research, we have been developing a systemic approach to introducing an assisted living platform for the home of older adults. To do so, we formed an interdisciplinary team that allows (1) to identify the user needs from a gerontological and psychological viewpoint; (2) to propose assistive applications designed by human factors and HCI experts, in collaboration with caregivers and users; (3) to develop and test applications designed and developed by software engineers; (4) to conduct a field study to assess the benefits of the platform and assistive applications, in collaboration with caregivers, by deploying the system at the actual homes.

Our research activities for assistive computing in the home are conducted under the *HomeAssist* project. This work takes the form of a platform offering an online catalog of assistive applications that orchestrate an open-ended set of networked objects. Our platform leverages DiaSuite to quickly and safely develop applications at a high level.

Our scientific achievements include the design principles of our platform, its key features to effectively assist individuals in their home, field studies to validate HomeAssist, the expansion of HomeAssist to serve individuals with ID, and the technology transfer of HomeAssist. Note that a complete presentation of this work, from a Cognitive Science perspective, is given in the doctoral thesis of Lucile Dupuy published last year (2016).

4.2.1. *Project-team positioning*

There is a range of platforms for assisted living aimed at older adults that have been developed for more than a decade. Most of these platforms are used in a setting where participants come to a research apartment to perform certain tasks. This setting makes it difficult to assess user acceptance and satisfaction of the proposed approaches because the user does not interact with the technology on a daily basis, over a period of time. Furthermore, older adults adopt routines to optimize their daily functioning at home. This situation calls for field studies in a naturalistic setting to strengthen the evaluation of assisted living platforms.

HomeAssist innovates in that it supports independent living across the activities of daily living and is validated by field studies in naturalistic setting.

4.3. Assistive computing on-the-go

Participants: Cécile Mazon, Léo Mendiboure, Hélène Sauzéon, Charles Consel.

We conduct research on assistive computing supported by mobile devices such as smart phones and tablets. Both research projects presented in this section are supported by tablets and leverage their functionalities to guide users with cognitive challenges performing activities and tasks, whether in mainstream schools to support inclusion or in residential settings to support their autonomy. The mobile nature of tablets allows to envision such devices as supporting users with cognitive challenges across a range of environments.

Many research projects bring cognitive-support applications to users based on tablets and smartphones. However, few projects equip users with such devices in actual mainstream environments, including stakeholders in the design process and targeting an autonomous usage of assistive applications. An additional originality of our approach is our interdisciplinary approach that allows us to integrate key psychological dimensions in our design, such as self-determination.

4.4. Life Plan

Participants: Audrey Landuran, Gregory Lecouvey, Léo Mendiboure, Quentin Barlas, Bernard N'Kaoua.

Elaboration of life plan is a major stage in the developmental trajectory of people with intellectual deficiency and is based on several capabilities : make choices, express them, be aware of their consequences, etc. However, these various capabilities can raise problems for people with Down syndrome. The aim of this project has been to design and validate a digital assistant that allows to help the individuals to make choices, to plan for the future and to define their life plan. The user centered methodology employed has allowed to involve the people with Down syndrome, their families and their caregivers in all stages of design. Different validations achieved attest for the accessible, friendly, funny character of the assistant and its ability to promote the expression of the life plan in accordance with individuals' choices, wishes and desires.

The digital assistant is now available online at : http://www.monprojetdevie.trisomie21-france.org/.

Figure 2. Sample screenshots from the Life Plan website.

<div align="center">

POLARIS Team

</div>

4. Application Domains

4.1. Large Computing Infrastructures

Supercomputers typically comprise thousands to millions of multi-core CPUs with GPU accelerators interconnected by complex interconnection networks that are typically structured as an intricate hierarchy of network switches. Capacity planning and management of such systems not only raises challenges in term of computing efficiency but also in term of energy consumption. Most legacy (SPMD) applications struggle to benefit from such infrastructure since the slightest failure or load imbalance immediately causes the whole program to stop or at best to waste resources. To scale and handle the stochastic nature of resources, these applications have to rely on dynamic runtimes that schedule computations and communications in an opportunistic way. Such evolution raises challenges not only in terms of programming but also in terms of observation (complexity and dynamicity prevents experiment reproducibility, intrusiveness hinders large scale data collection, ...) and analysis (dynamic and flexible application structures make classical visualization and simulation techniques totally ineffective and require to build on *ad hoc* information on the application structure).

4.2. Next-Generation Wireless Networks

Considerable interest has arisen from the seminal prediction that the use of multiple-input, multiple-output (MIMO) technologies can lead to substantial gains in information throughput in wireless communications, especially when used at a massive level. In particular, by employing multiple inexpensive service antennas, it is possible to exploit spatial multiplexing in the transmission and reception of radio signals, the only physical limit being the number of antennas that can be deployed on a portable device. As a result, the wireless medium can accommodate greater volumes of data traffic without requiring the reallocation (and subsequent re-regulation) of additional frequency bands. In this context, throughput maximization in the presence of interference by neighboring transmitters leads to games with convex action sets (covariance matrices with trace constraints) and individually concave utility functions (each user's Shannon throughput); developing efficient and distributed optimization protocols for such systems is one of the core objectives of Theme 5.

Another major challenge that occurs here is due to the fact that the efficient physical layer optimization of wireless networks relies on perfect (or close to perfect) channel state information (CSI), on both the uplink and the downlink. Due to the vastly increased computational overhead of this feedback – especially in decentralized, small-cell environments – the ongoing transition to fifth generation (5G) wireless networks is expected to go hand-in-hand with distributed learning and optimization methods that can operate reliably in feedback-starved environments. Accordingly, one of POLARIS' application-driven goals will be to leverage the algorithmic output of Theme 5 into a highly adaptive resource allocation framework for next-gneration wireless systems that can effectively "learn in the dark", without requiring crippling amounts of feedback.

4.3. Energy and Transportation

Participant: Nicolas Gast.

This work is mainly done within the Quanticol European project.

Smart urban transport systems and smart grids are two examples of collective adaptive systems. They consist of a large number of heterogeneous entities with decentralised control and varying degrees of complex autonomous behaviour. Within the QUANTICOL project, we develop an analysis tools to help to reason about such systems. Our work relies on tools from fluid and mean-field approximation to build decentralized algorithms that solve complex optimization problems. We focus on two problems: decentralized control of electric grids and capacity planning in vehicle-sharing systems to improve load balancing.

RAP2 Team (section vide)

REGAL Project-Team (section vide)

<p style="text-align:center">**RMOD Project-Team**</p>

4. Application Domains

4.1. Programming Languages and Tools

Many of the results of RMoD are improving programming languages or development tools for such languages. As such the application domain of these results is as varied as the use of programming languages in general. Pharo, the language that RMoD develops, is used for a very broad range of applications. From pure research experiments to real world industrial use (the Pharo Consortium, http://consortium.pharo.org, has more than 25 company members).

Examples are web applications, server backends for mobile applications or even graphical tools and embedded applications

4.2. Software Reengineering

Moose is a language-independent environment for reverse and re-engineering complex software systems. Moose provides a set of services including a common meta-model, metrics evaluation and visualization. As such Moose is used for analyzing software systems to support understanding and continuous development as well as software quality analysis.

<p style="text-align:center;color:red;font-weight:bold;">ROMA Project-Team</p>

4. Application Domains

4.1. Applications of sparse direct solvers

Sparse direct (e.g., multifrontal solvers that we develop) solvers have a wide range of applications as they are used at the heart of many numerical methods in computational science: whether a model uses finite elements or finite differences, or requires the optimization of a complex linear or nonlinear function, one often ends up solving a system of linear equations involving sparse matrices. There are therefore a number of application fields, among which some of the ones cited by the users of our sparse direct solver MUMPS are: structural mechanics, seismic modeling, biomechanics, medical image processing, tomography, geophysics, electromagnetism, fluid dynamics, econometric models, oil reservoir simulation, magneto-hydro-dynamics, chemistry, acoustics, glaciology, astrophysics, circuit simulation, and work on hybrid direct-iterative methods.

SOCRATE Project-Team (section vide)

SPIRALS Project-Team

4. Application Domains

4.1. Introduction

Although our research is general enough to be applied to many application domains, we currently focus on applications and distributed services for the retail industry and for the digital home. These two application domains are supported by a strong expertise in mobile computing and in cloud computing that are the two main target environments on which our research prototypes are built, for which we are recognized, and for which we have already established strong collaborations with the industrial ecosystem.

4.2. Distributed software services for the retail industry

This application domain is developed in relation with the PICOM (*Pôle de compétivité Industries du Commerce*) cluster. We have established strong collaborations with local companies in the context of former funded projects, such as Cappucino and Macchiato, which focused on the development of a new generation of mobile computing platforms for e-commerce. We are also involved in the Datalyse and OCCIware funded projects that define cloud computing environments with applications for the retail industry. Finally, our activities in terms of crowd-sensing and data gathering on mobile devices with the APISENSE$^{®}$ platform share also applications for the retail industry.

4.3. Distributed software services for the digital home

We are developing new middleware solutions for the digital home, in particular through our long standing collaboration with Orange Labs. We are especially interested in developing energy management and saving solutions with the POWERAPI software library for distributed environments such the ones that equip digital homes. We are also working to bridge the gap between distributed services hosted on home gateways and distributed services hosted on the cloud to be able to smoothly transition between both environments. This work is especially conducted with the SALOON platform.

STORM Project-Team

4. Application Domains

4.1. Application Domains

The application of our work concerns linear algebra, solvers and fast-multipole methods, in collaboration with other Inria teams and with industry. This allows a wide range of scientific and industrial applications possibly interested in the techniques we propose, in the domain of high performance computing but also in order to compute intensive embedded applications. In terms of direct application, the software developed in the team are used in applications in various fields, ranging from seismic, mechanic of fluids, molecular dynamics, high energy physics or material simulations. Similarly, the domains of image processing and signal processing can take advantage of the expertise and software of the team.

<p style="color:red; text-align:center; font-weight:bold;">TACOMA Team</p>

4. Application Domains

4.1. Pervasive applications in Smart Building

A Smart Building is a living space equipped with information-and-communication-technology (ICT) devices conceived to collaborate in order to anticipate and respond to the needs of the occupants, working to promote their comfort, convenience, security and entertainment while preserving their natural interaction with the environment.

The idea of using the Pervasive Computing paradigm in the Smart Building domain is not new. However, the state-of-the-art solutions only partially adhere to its principles. Often the adopted approach consists in a heavy deployment of sensor nodes, which continuously send a lot of data to a central elaboration unit, in charge of the difficult task of extrapolating meaningful information using complex techniques. This is a *logical approach*. TACOMA proposed instead the adoption of a *physical approach*, in which the information is spread in the environment, carried by the entities themselves, and the elaboration is directly executed by these entities "inside" the physical space. This allows performing meaningful exchanges of data that will thereafter need a less complicated processing compared to the current solutions. The result is a smart environment that can, in an easier and better way, integrate the context in its functioning and thus seamlessly deliver more useful and effective user services. Our contribution aims at implementing the physical approach in a smarter environment, showing a solution for improving both comfort and energy savings.

4.2. Metamorphic House

The motivation for metamorphic houses is that many countries, including France, are going through socio-demographic evolutions, like growth of life expectancy and consequent increase in the number of elderly people, urbanization and resource scarcity. Households experience financial restrictions, while housing costs increase with the raise of real estate and energy prices [6].

Important questions arise concerning the future of housing policies and ways of living. We observe novel initiatives like participative housing and developing behaviors, including house-sharing, teleworking and longer stay of children in parents' homes.

To tackle the challenges raised by these emerging phenomena, future homes will have to be modular, upgradeable, comfortable, sparing of resources. They should be integrated in the urban context and exchange information with other homes, contribute to reducing the distances to be covered daily and respect the characteristics of the territory where they are located.

To reach these goals, metamorphic domestic environments will modify their shape and behavior to support activities and changes in life cycle of occupants, increase comfort and optimize the use of resources. Thanks to Information and Communication Technologies (ICT) and adaptive building elements, the same physical spaces will be transformed for different uses, giving inhabitants the illusion of living in bigger, more adapted and more comfortable places.

4.3. Automation in Smart City

The domain of Smart Cities is still young but it is already a huge market which attract number of companies and researchers. It is also multi-fold as the words "smart city" gather multiple meanings. Among them one of the main responsibilities of a city, is to organize the transportation of goods and people. In intelligent transportation systems (ITS), ICT technologies have been involved to improve planification and more generally efficiency of journeys within the city. We are interested in the next step where efficiency would be improved locally relying on local interactions between vehicles, infrastructure and people (smartphones).

For the future autonomous vehicle are now in the spotlight, since a lot of works has been done in recent years in automotive industry as well as in academic research centers. Such unmanned vehicle could strongly impact the organisation of the transportation in our cities. However, due to the lack of a definition of what is an "autonomous" vehicle it remains still difficult to see how these vehicles will interact with their environment (eg. road, smart city, houses, grid, etc.). From augmented perception to fully cooperative automated vehicle, the autonomy covers various realities in terms of interaction the vehicle relies on. The extended perception relies on communication between the vehicle and surrounding roadside equipments. This help the driving system to build and maintain an accurate view of the environment. But at this first stage the vehicle only uses its own perception to make its decisions. At a second stage, it will take advantages of local interaction with other vehicles through car-to-car communications to elaborate a better view of its environment. Such "cooperative autonomy" does not try to reproduce the human behavior anymore, it strongly relies on communication between vehicles and/or with the infrastructure to make decision and to acquire information on the environment. Part of the decision could be centralized (almost everything for an automatic metro) or coordinated by a roadside component. The decision making could even be fully distributed but this put high constraints on the communications. Automated vehicles are just an exemple of smart city automated processes that will have to share information within the surrounding to make their decisions.

4.4. Pervasive applications in uncontrolled environnements

Some limitations of existing RFID technology become challenging: unlike standard RFID application scenarios, pervasive computing often involves uncontrolled environment for RFID, where tags and reader have to operate in much more difficult situations that those usually encountered or expected for classical RFID systems.

RFID technology is to avoid missing tags when reading multiple objects, as reading reliability is affected by various effects such shadowing or wave power absorption by some materials. The usual applications of RFID operate in a controlled environment in order to reduce the risk of missing tags while scanning objects.

In pervasive computing applications, a controlled reading environment is extremely difficult to achieve, as one of the principle is to enhance existing processes "in situ", unlike the controlled conditions that can be found in industrial processes. Consider for example a logistic application, where RFID tags could be used on items inside a package in order to check for its integrity along the shipping process. Tags would likely be placed randomly on items inside the package, and reading conditions would be variable depending on where the package is checked.

RFID operation in uncontrolled environments is challenging because RFID performance is affected by multiple parameters, in particular:

- Objects materials (on which tags are attached to),
- Materials in the surrounding environment,
- RFID frequency spectrum,
- Antenna nature and placement with respect to the tags.

In controlled environment, the difficulty to read tags can be limited by using the appropriate parameters to maximize the RFID performance for the application. But in many cases, it is needed to read large number of objects of various nature, arranged randomly in a given area or container. **Most pervasive computing applications fall in this context**.

<p style="text-align:center">**TADaaM Project-Team**</p>

4. Application Domains

4.1. Mesh-based applications

TADAAM targets scientific simulation applications on large-scale systems, as these applications present huge challenges in terms of performance, locality, scalability, parallelism and data management. Many of these HPC applications use meshes as the basic model for their computation. For instance, PDE-based simulations using finite differences, finite volumes, or finite elements methods operate on meshes that describe the geometry and the physical properties of the simulated objects. This is the case for at least two thirds of the applications selected in the 9th PRACE. call [0], which concern quantum mechanics, fluid mechanics, climate, material physic, electromagnetism, etc.

Mesh-based applications not only represent the majority of HPC applications running on existing supercomputing systems, yet also feature properties that should be taken into account to achieve scalability and performance on future large-scale systems. These properties are the following:

Size   Datasets are large: some meshes comprise hundreds of millions of elements, or even billions.

Dynamicity   In many simulations, meshes are refined or coarsened at each time step, so as to account for the evolution of the physical simulation (moving parts, shockwaves, structural changes in the model resulting from collisions between mesh parts, etc.).

Structure   Many meshes are unstructured, and require advanced data structures so as to manage irregularity in data storage.

Topology   Due to their rooting in the physical world, meshes exhibit interesting topological properties (low dimensionality embedding, small maximum degree, large diameter, etc.). It is very important to take advantage of these properties when laying out mesh data on systems where communication locality matters.

All these features make mesh-based applications a very interesting and challenging use-case for the research we want to carry out in this project. Moreover, we believe that our proposed approach and solutions will contribute to enhance these applications and allow them to achieve the best possible usage of the available resources of future high-end systems.

---

[0]http://www.prace-ri.eu/prace-9th-regular-call/

WHISPER Project-Team

4. Application Domains

4.1. Linux

Linux is an open-source operating system that is used in settings ranging from embedded systems to supercomputers. The most recent release of the Linux kernel, v4.14, comprises over 16 million lines of code, and supports 30 different families of CPU architectures, around 50 file systems, and thousands of device drivers. Linux is also in a rapid stage of development, with new versions being released roughly every 2.5 months. Recent versions have each incorporated around 13,500 commits, from around 1500 developers. These developers have a wide range of expertise, with some providing hundreds of patches per release, while others have contributed only one. Overall, the Linux kernel is critical software, but software in which the quality of the developed source code is highly variable. These features, combined with the fact that the Linux community is open to contributions and to the use of tools, make the Linux kernel an attractive target for software researchers. Tools that result from research can be directly integrated into the development of real software, where it can have a high, visible impact.

Starting from the work of Engler et al. [41], numerous research tools have been applied to the Linux kernel, typically for finding bugs [39], [58], [71], [81] or for computing software metrics [47], [87]. In our work, we have studied generic C bugs in Linux code [9], bugs in function protocol usage [52], [53], issues related to the processing of bug reports [75] and crash dumps [46], and the problem of backporting [70], [82], illustrating the variety of issues that can be explored on this code base. Unique among research groups working in this area, we have furthermore developed numerous contacts in the Linux developer community. These contacts provide insights into the problems actually faced by developers and serve as a means of validating the practical relevance of our work.

4.2. Device Drivers

Device drivers are essential to modern computing, to provide applications with access, via the operating system, to physical devices such as keyboards, disks, networks, and cameras. Development of new computing paradigms, such as the internet of things, is hampered because device driver development is challenging and error-prone, requiring a high level of expertise in both the targeted OS and the specific device. Furthermore, implementing just one driver is often not sufficient; today's computing landscape is characterized by a number of OSes, *e.g.*, Linux, Windows, MacOS, BSD and many real time OSes, and each is found in a wide range of variants and versions. All of these factors make the development, porting, backporting, and maintenance of device drivers a critical problem for device manufacturers, industry that requires specific devices, and even for ordinary users.

The last fifteen years have seen a number of approaches directed towards easing device driver development. Réveillère, who was supervised by G. Muller, proposes Devil [7], a domain-specific language for describing the low-level interface of a device. Chipounov *et al.* propose RevNic, [33] a template-based approach for porting device drivers from one OS to another. Ryzhyk *et al.* propose Termite, [72], [73] an approach for synthesizing device driver code from a specification of an OS and a device. Currently, these approaches have been successfully applied to only a small number of toy drivers. Indeed, Kadav and Swift [49] observe that these approaches make assumptions that are not satisfied by many drivers; for example, the assumption that a driver involves little computation other than the direct interaction between the OS and the device. At the same time, a number of tools have been developed for finding bugs in driver code. These tools include SDV [23], Coverity [41], CP-Miner, [57] PR-Miner [58], and Coccinelle [8]. These approaches, however, focus on analyzing existing code, and do not provide guidelines on structuring drivers.

In summary, there is still a need for a methodology that first helps the developer understand the software architecture of drivers for commonly used operating systems, and then provides tools for the maintenance of existing drivers.

<p style="text-align:center; color:red">ALICE Project-Team</p>

4. Application Domains

4.1. Geometric Tools for Simulating Physics with a Computer

Numerical simulation is the main targeted application domain for the geometry processing tools that we develop. Our mesh generation tools are tested and evaluated within the context of our cooperation with the Gocad consortium, with applications in oil exploration and geomechanics, through co-advised Ph.D. theses (Arnaud Botella, Julien Renaudeau). We think that the hex-dominant meshes that we generate have geometrical properties that make them suitable for some finite element analyses. We work on evaluating and measuring their impact with simple problems (heat equation, linear elasticity) and then practical applications (unfolding geological layer), with the Ph.D. thesis of Maxence Reberol.

In numerical simulation, developing discrete formulations that satisfy the conservation laws (conservation of mass, conservation of energy, conservation of momentum) is important to ensure that the numerical simulation faithfully reflects the behavior of the physics. There are interesting relations with optimal transport theory, as explained by Benamou and Brenier who developed a numerical algorithm for optimal transport that uses a fluid dynamics formulation [20]. Conversely, some dynamics can be approximated by a series of optimal transport problems, as in the Jordan-Kinderlehrer-Otto scheme [23] and in recent works by Mérigot. We started developing efficient geometric algorithms and optimisation methods that may serve as the basis for implementing these numerical methods in 3D. We started discussions / cooperation projects with Quentin Mérigot (MOKAPLAN project).

4.2. Fabrication

Our work around fabrication and additive manufacturing finds applications in different fields. Our algorithms for fast geometric computations on solids (boolean operations, morphological operations) are useful to model a variety of shapes, from mechanical engineering parts to prosthetics for medical applications.

Our by-example techniques allow for simpler modeling and processing of very intricate geometries and therefore also find applications in art and design, for unusual shapes that would be very difficult to obtain otherwise. Extensions of these techniques also find applications for reproducing naturally occurring micro-structures from a scanned sample.

<p align="center" style="color:red">ALMANACH Team</p>

4. Application Domains

4.1. Application domains of NLP and Computational Humanities

ALMAnaCH's research areas cover Natural Language Processing (nowadays recognised as a sub-domain of Artificial Intelligence) and Digital Humanities. Application domains are therefore numerous, as witnessed by ALMAnaCH's multiple academic and industrial collaborations, for which see the relevant sections. Examples of application domains include:

- Information extraction, information retreival, text mining (ex.: opinion surveys)
- Text generation, text simplification, automatic summarisation
- Spelling correction (writing aid, post-OCR, normalisation of noisy/non-canonical texts)
- Machine translation, computer-aided translation
- Chatbots, conversational agents, question answering systems
- Medical applications (early diagnosis, language-based medical monitoring...)
- Applications in linguistics (modelling languages and their evolution, sociolinguistic studies...)
- Digital humanities (exploitation of text documents, for instance in historical research)

AVIZ Project-Team (section vide)

CEDAR Team

4. Application Domains

4.1. Cloud Computing

Cloud computing services are strongly developing and more and more companies and institutions resort to running their computations in the cloud, in order to avoid the hassle of running their own infrastructure. Today's cloud service providers guarantee machine availabilities in their Service Level Agreement (SLA), without any guarantees on performance measures according to a specific cost budget. Running analytics on big data systems require the user not to only reserve the suitable cloud instances over which the big data system will be running, but also setting many system parameters like the degree of parallelism and granularity of scheduling. Chosing values for these parameters, and chosing cloud instances need to meet user objectives regarding latency, throughput and cost measures, which is a complex task if it's done manually by the user. Hence, we need need to transform cloud service models from availabily to user performance objective rises and leads to the problem of multi-objective optimization. Research carried out in the team within the ERC project "Big and Fast Data Analytics" aims to develop a novel optimization framework for providing guarantees on the performance while controlling the cost of data processing in the cloud.

4.2. Computational Journalism

Modern journalism increasingly relies on content management technologies in order to represent, store, and query source data and media objects themselves. Writing news articles increasingly requires consulting several sources, interpreting their findings in context, and crossing links between related sources of information. CEDARresearch results directly applicable to this area provide techniques and tools for rich Web content warehouse management. Within the ANR ContentCheck project, and also as part of our international collaboration with the AIST institute from Japan, we work on one hand, to lay down foundations for computational data journalism and fact checking, and also work to devise concrete algorithms and platforms to help journalists perform their work better and/or faster. This work is carried in collaboration with Le Monde's "Les Décodeurs".

On a related topic, heterogeneous data integration under a virtual graph abstract model is studied within the ICODA Inria project which has started in September 2017. There, we collaborate with Les Décodeurs as well as with Ouest France and Agence France Presse (AFP). The data and knowledge integration framework resulting from this work will support journalists' effort to organize and analyze their knowledge and exploit it in order to produce new content.

4.3. Open Data Intelligence

The Web is a vast source of information, to which more is added every day either in unstructured form (Web pages) or, increasingly, as partially structured sources of information, in particular as Open Data sets, which can be seen as connected graphs of data, most frequently described in the RDF data format recommended by the W3C. Further, RDF data is also the most appropriate format for representing structured information extracted automatically from Web pages, such as the DBPedia database extracted from Wikipedia or Google's InfoBoxes. We work on this topic within the 4-year project ODIN started in 2014.

4.4. Genomics

One particular case of area where the increase in data production is the more consequent is genomic data, indeed the amount of data produced doubles every 7 months. Thus we want to bring the expertise from the database and big data community to help both scale the existing algorithms and design new algorithms that are scalable from the ground up.

CHROMA Project-Team

4. Application Domains

4.1. Introduction

Applications in Chroma are organized in two main domains : **i) Future cars and transportation systems and ii) Services robotics.** These domains correspond to the experimental fields initiated in Grenoble (eMotion team) and in Lyon (CITI lab). However, the scientific objectives described in the previous sections are intended to apply equally to both applicative domains. Even our work on Bayesian Perception is today applied to the intelligent vehicle domain, we aim to generalize to any mobile robots. The same remark applies to the work on multi-agent decision making. We aim to apply algorithms to any fleet of mobile robots (service robots, connected vehicles, UAVs). This is the philosophy of the team since its creation.

Figure 5. Most of the Chroma platforms: the Pepper robot, a fleet of (22) Turtlebot 2, one of the 4 Bebop drones and the equipped Toyota Lexus.

4.2. Future cars and transportation systems

Thanks to the introduction of new sensor and ICT technologies in cars and in mass transportation systems, and also to the pressure of economical and security requirements of our modern society, this application domain is quickly changing. Various technologies are currently developed by both research and industrial laboratories. These technologies are progressively arriving at maturity, as it is witnessed by the results of large scale experiments and challenges such as the Google's car project and several future products announcements made by the car industry. Moreover, the legal issue starts to be addressed in USA (see for instance the recent laws in Nevada and in California authorizing autonomous vehicles on roads) and in several other countries (including France).

In this context, we are interested in the development of ADAS [0] systems aimed at improving comfort and safety of the cars users (e.g., ACC, emergency braking, danger warnings), and of Fully Autonomous Driving functions for controlling the displacements of private or public vehicles in some particular driving situations and/or in some equipped areas (e.g., automated car parks or captive fleets in downtown centers or private sites).

---

[0]Advanced Driver Assistance Systems

Since about 8 years, we are collaborating with Toyota and with Renault-Nissan on these applications (bilateral contracts, PhD Theses, shared patents), but also recently with Volvo group (PhD thesis started in 2016). We are also strongly involved (since 2012) in the innovation project Perfect of the IRT [0] Nanoelec (transportation domain). In 2016, we have been awarded a European H2020 ECSEL project [0] involving major European automotive constructors and car suppliers. In this project, Chroma is focusing on the embedded perception component (models and algorithms, including the certification issue), in collaboration with Renault, Valeo and also with the Inria team TAMIS (Rennes). Chroma is also involved in the ANR project "Valet" (2015-2018) coordinated by the Inria team RITS (Rocquencourt), dealing with automatic redistribution of car-sharing vehicles and parking valet; Chroma is involved in the pedestrian-vehicle interaction for a safe navigation.

In this context, Chroma has two experimental vehicles equipped with various sensors (a Toyota Lexus and a Renault Zoe, see. Fig. 5 and Fig. 2), which are maintained by Inria-SED [0] and that allow the team to perform experiments in realistic traffic conditions (Urban, road and highway environments). The Zoe car has been automated in December 2016, through our collaboration with the team of P. Martinet (IRCCyN Lab, Nantes), that allow new experiments in the team.

4.3. Services robotics

Service robotics is an application domain quickly emerging, and more and more industrial companies (e.g., IS-Robotics, Samsung, LG) are now commercializing service and intervention robotics products such as vacuum cleaner robots, drones for civil or military applications, entertainment robots ... One of the main challenges is to propose robots which are sufficiently robust and autonomous, easily usable by non-specialists, and marked at a reasonable cost. We are involved in developing observation and surveillance systems, by using ground robots (Turtlebot fleet) or aerial ones (ANR VIMAD [0]), see Fig. 5 . Since 2016, we develop solutions for 3D observation/exploration of complex scenes or environments with a fleet of UAVs (Inria ADT CORDES) or ground robots (COMODYS FIL project).

A more recent challenge for the coming decade is to develop robotized systems for assisting elderly and/or disabled people. In the continuity of our work in the IPL PAL [0], we aim to propose smart technologies to assist electric wheelchair users in their displacements (see Figure 2 for illustration). We address the problem of assisting the user for joining a group of people and navigating in crowded environments, in cooperation with Inria Lagadic team (Rennes) and also in our recent ANR Hianic. Another emerging application to assist people is telepresence robot. In 2016 we started the TENSIVE project, funded by the Region, with the team of G. Bailly from GIPSA Lab (Grenoble) and with the Awabot and Hoomano companies (in Lyon). The project aims to improve the driving of such robots by providing a social and autonomous navigation (PhD of R. Cambuzat). Moreover, the project is supported by INSA-CITI Lab. through the acquisition of a Pepper robot (see Fig. 5).

---

[0] Institut de Recherche Technologique
[0] ENABLE-S3: European Initiative to Enable Validation for Highly Automated Safe and Secure Systems.
[0] Service Expérimentation et Développement
[0] Navigation autonome des drones aériens avec la fusion des données visuelles et inertielles, lead by A. Martinelli, Chroma.
[0] Personnaly assisted Living

COML Team

4. Application Domains

4.1. Speech processing for underresourced languages

We plan to apply our algorithms for the unsupervised discovery of speech units to problems relevant to language documentation and the construction of speech processing pipelines for underresourced languages.

4.2. Tools for the analysis of naturalistic speech corpora

Daylong recordings of speech in the wild gives rise a to number of specific analysis difficulties. We plan to use our expertise in speech processing to develop tools for performing signal processing and helping annotation of such resources for the purpose of phonetic or linguistic analysis.

DEFROST Project-Team

4. Application Domains

4.1. Industry

Robotics in the manufacturing industry is already highly diffused and is one of the ways put forward to maintain the level of competitiveness of companies based in France and to avoid relocation in cheap labor countries. Yet, in France, it is considered that the level of robotization is insufficient compared to Germany, for instance. One of the challenge is the high investment cost for buying robotic arms. In the recent years, it has led the development of « generic » and « flexible » (but rigid) robotic solution that can be produced in series. But their applicability to specific tasks is still challenging or too costly. With the development of 3D printing, we can imagine the development of a complete opposite strategy: a « task-specific » design of robots. Given a task that needs to be performed by a deformable robot: we would optimize the shape of its structure to create the set of desired motion. An second important aspect is the reduction of the manufacturing cost: It is often anticipated that the cost of deformable robots will be low compared to classical rigid robotics. The robot could be built on one piece using rapid prototyping or 3D printers and be more adapted for collaborative work with operators. In this area, using soft materials are particularly convenient as they provide a mass/carried load ratio several orders higher than traditional robots, highly decreasing the kinetic energy and so increasing the motion speed allowed in presence of humans. Moreover, the technology allows more efficient and ergonomic wearable robotic devices, opening the options for exo-skeletons. This remains to be put in place, but it can open new perspectives in robotic applications. A last remarkable property of soft robots is their adaptability to fragile or tortuous environment. For some particular industry (chemistry, food industry...) this could also be an advantage compared to existing rigid solutions. For instance, the German company https://www.festo.com/group/en/cms/12747.htm, key player in the industrial robots field, is experiencing with deformable trunk robot and we are working on their accurate control.

This year (2017), we have started to work more concretely on this domain, with a research contrat with TDR company on design of deformable graspers.

4.2. Personal and service robotics

The personal and service robotics are considered as an important source of economic expansion in the coming years. The potential applications are numerous and particularly include the challenge of finding robotic solutions for active and healthy aging at home. We plan to develop functional orthosis for which it is better not to have a rigid exoskeleton that is particularly not comfortable. These orthosis will be ideally personalised for each patient and built using rapid prototyping. On this topic, the place of our team will be to provide algorithms for controlling the robots. We will find some partners to build these robots that would fall in the category of « wearable robots ». With this thematic we also connect with a strong pole of excellence of the region on intelligent textile (see http://www.up-tex.fr/Up-Tex) and with the strategic plan of Inria (Improving Rehabilitation and Autonomy).

4.3. Entertainment industry and arts

Robots have a long history with entertainment and arts where https://en.wikipedia.org/wiki/Animatronicsanimatronics have been used since years for cinematographic shootings, theater, amusement parc (https://en.wikipedia.org/wiki/List_of_Disney_attractions_using_Audio-AnimatronicsDisney's audio-animatronic) and performing arts. We believe that soft robots could be a good support for art. We have currently a collaboration with the artist Jonathan Pepe (see http://jonathanpepe.tumblr.com/Haruspices).

Figure 1. Exobiote project

EX-SITU Project-Team

4. Application Domains

4.1. Creative industries

We work closely with creative professionals in the arts and in design, including music composers, musicians, and sound engineers; painters and illustrators; dancers and choreographers; theater groups; graphic and industrial designers; and architects.

4.2. Scientific research

We work with creative professionals in the sciences and engineering, including neuroscientists and doctors; programmers and statisticians; chemists and astrophysicists; and researchers in fluid mechanics.

4. Application Domains

4.1. Application Domains

Cognitive Sciences The computational modelling of life-long learning and development mechanisms achieved in the team centrally targets to contribute to our understanding of the processes of sensorimotor, cognitive and social development in humans. In particular, it provides a methodological basis to analyze the dynamics of the interaction across learning and inference processes, embodiment and the social environment, allowing to formalize precise hypotheses and later on test them in experimental paradigms with animals and humans. A paradigmatic example of this activity is the Neurocuriosity project achieved in collaboration with the cognitive neuroscience lab of Jacqueline Gottlieb, where theoretical models of the mechanisms of information seeking, active learning and spontaneous exploration have been developed in coordination with experimental evidence and investigation, see https://flowers.inria.fr/neurocuriosityproject/.

Personal and lifelong learning robotics Many indicators show that the arrival of personal robots in homes and everyday life will be a major fact of the 21st century. These robots will range from purely entertainment or educative applications to social companions that many argue will be of crucial help in our society. Yet, to realize this vision, important obstacles need to be overcome: these robots will have to evolve in unpredictable homes and learn new skills in a lifelong manner while interacting with non-engineer humans after they left factories, which is out of reach of current technology. In this context, the refoundation of intelligent systems that developmental robotics is exploring opens potentially novel horizons to solve these problems. In particular, this application domain requires advances in artificial intelligence that go beyond the current state-of-the-art in fields like deep learning. Currently these techniques require tremendous amounts of data in order to function properly, and they are severaly limited in terms of incremental and transfer learning. One of our goals is to drastically reduce the amount of data required in order for this very potent field to work. We try to achieve this by making neural networks aware of their knowledge, i.e. we introduce the concept of uncertainty, and use it as part of intrinsically motivated multitask learning architectures, and combined with techniques of learning by imitation.

Human-Robot Collaboration. Robots play a vital role for industry and ensure the efficient and competitive production of a wide range of goods. They replace humans in many tasks which otherwise would be too difficult, too dangerous, or too expensive to perform. However, the new needs and desires of the society call for manufacturing system centered around personalized products and small series productions. Human-robot collaboration could widen the use of robot in this new situations if robots become cheaper, easier to program and safe to interact with. The most relevant systems for such applications would follow an expert worker and works with (some) autonomy, but being always under supervision of the human and acts based on its task models.

Environment perception in intelligent vehicles. When working in simulated traffic environments, elements of FLOWERS research can be applied to the autonomous acquisition of increasingly abstract representations of both traffic objects and traffic scenes. In particular, the object classes of vehicles and pedestrians are if interest when considering detection tasks in safety systems, as well as scene categories ("scene context") that have a strong impact on the occurrence of these object classes. As already indicated by several investigations in the field, results from present-day simulation technology can be transferred to the real world with little impact on performance. Therefore, applications of FLOWERS research that is suitably verified by real-world benchmarks has direct applicability in safety-system products for intelligent vehicles.

Automated Tutoring Systems. Optimal teaching and efficient teaching/learning environments can be applied to aid teaching in schools aiming both at increase the achievement levels and the reduce time needed. From a practical perspective, improved models could be saving millions of hours of students' time (and effort) in learning. These models should also predict the achievement levels of students in order to influence teaching practices.

GRAPHDECO Project-Team (section vide)

<p align="center">**GRAPHIK Project-Team**</p>

4. Application Domains

4.1. Agronomy

Agronomy is a strong expertise domain in the area of Montpellier. Some INRA researchers (computer scientists) are members of GraphIK, and more generally we closely collaborate with the Montpellier research laboratory IATE, a join unit of INRA and other organisms. A major issue for INRA is modeling agrifood chains (i.e., the chain of all processes leading from the plants to the final products, including waste treatment). This modeling has several objectives. It provides better understanding of the processes from begin to end, which aids in decision making, with the aim of improving the quality of the products and decreasing the environmental impact. It also facilitates knowledge sharing between researchers, as well as the capitalization of expert knowledge and "know how". This last point is particularly important in areas strongly related to a "terroir" (like in cheese or wine making), where knowledge and "know how" are transmitted by experience, with the risk of non-sustainability of the specific skills. For all these reasons, INRA became very interested in developing knowledge engineering methods. An agrifood chain analysis is a highly complex procedure since it relies on numerous criteria of various types: environmental, economical, functional, sanitary, etc. Quality objectives imply different stakeholders, technicians, managers, professional organizations, end-users, public organizations, etc. Since the goals of the implied stakeholders may be divergent, decision making raises arbitration issues. In this context, our first investigations led to identify decision support based on argumentation frameworks as a promising topic, as well as the representation and processing of preferences.

4.2. Data Journalism

One of today's major issues in data science is to design techniques and algorithms that allow analysts to efficiently infer useful information and knowledge by inspecting heterogeneous information sources, from structured data to unstructured content. We take data journalism as an emblematic use-case, which stands at the crossroad of multiple research fields: content analysis, data management, knowledge representation and reasoning, visualization and human-machine interaction. We are particularly interested in issues raised by the design of data and knowledge management systems that will support data journalism. These systems include an ontology that typically expresses domain knowledge, heterogeneous data sources, and mappings that relate these data sources expressed with their own vocabulary and querying capabilities, to a (possibly virtual) factbase expressed using the ontological vocabulary. Ontologies play a central role as they act both as a mediation layer that glue together pieces of knowledge extracted from heterogeneous data sources, and as an inference layer that allow to draw new knowledge. In the context of data journalism, those ontologies require challenging features that we need to take into account:

- the wide range of topics addressed in journalism requires a rich top-level ontology, though very specific ontologies might be required to handle specific knowledge (e.g. detailed knowledge on finance to handle the panama papers).

- in data journalism, each piece of knowledge requires different timestamps (temporal information represented within the data, for instance when an event effectively takes place, and temporal information about the data itself, for instance when this event is recorded / validated in the system). Temporal relations (such as Allen's) can be used to express constraints between timestamps and ensure the consistency of the (virtual) knowledge base.

- in data journalism, each piece of knowledge has an identified source. The analysis of conflicting knowledge in the (virtual) knowledge base has to take the source fiability into account.

Besides pure knowledge representation and reasoning issues, querying such systems raise issues at the crossroad of data and knowledge management. In particular, the notion of mappings has to be revisited in the light of the reasoning capabilities enabled by the ontology. More generally, the consistency and the efficiency of the system cannot be ensured by considering the components of the system in isolation (i.e., the ontology, data sources and mappings), but require to study the interactions between these components and to consider the system as a whole.

HEPHAISTOS Project-Team

4. Application Domains

4.1. Domain

While the methods developed in the project can be used for a very broad set of application domains (for example we have an activity in CO2 emission allowances, it is clear that the size of the project does not allow us to address all of them. Hence we have decided to focus our applicative activities on *mechanism theory*, where we focus on *modeling*, *optimal design* and *analysis* of mechanisms. Along the same line our focus is *robotics* and especially *service robotics* which includes rescue robotics, rehabilitation and assistive robots for elderly and handicapped people. Although these topics were new for us when initiating the project we have spent two years determining priorities and guidelines by conducting about 200 interviews with field experts (end-users, praticians, family and caregivers, institutes), establishing strong collaboration with them (e.g. with the CHU of Nice-Cimiez) and putting together an appropriate experimental setup for testing our solutions. A direct consequence of setting up this research framework is a reduction in our publication and contract activities. But this may be considered as an investment as assistance robotics is a long term goal. It must be reminded that we are able to manage a large variety of problems in totally different domains only because interval analysis, game theory and symbolic tools provides us the methodological tools that allow us to address completely a given problem from the formulation and analysis up to the very final step of providing numerical solutions.

HYBRID Project-Team

4. Application Domains

4.1. Overview

The research program of Hybrid team aims at next generations of virtual reality and 3D user interfaces which could possibly address both the "body" and "mind" of the user. Novel interaction schemes are designed, for one or multiple users. We target better integrated systems and more compelling user experiences.

The applications of our research program correspond to the applications of virtual reality technologies which could benefit from the addition of novel body-based or mind-based interaction capabilities:

- **Industry:** with training systems, virtual prototyping, or scientific visualization;
- **Medicine:** with rehabilitation and reeducation systems, or surgical training simulators;
- **Entertainment:** with 3D web navigations, video games, or attractions in theme parks,
- **Construction:** with virtual mock-ups design and review, or historical/architectural visits.

ILDA Project-Team

4. Application Domains

4.1. Mission-critical systems

Mission-critical contexts of use include emergency response & management, and critical infrastructure operations, such as public transportation systems, communications and power distribution networks, or the operations of large scientific instruments such as particle accelerators and astronomical observatories. Central to these contexts of work is the notion of situation awareness [26], i.e., how workers perceive and understand elements of the environment with respect to time and space, such as maps and geolocated data feeds from the field, and how they form mental models that help them predict future states of those elements. One of the main challenges is how to best assist subject-matter experts in constructing correct mental models and making informed decisions, often under time pressure. This can be achieved by providing them with, or helping them efficiently identify and correlate, relevant and timely information extracted from large amounts of raw data, taking into account the often cooperative nature of their work and the need for task coordination. With this application area, our goal is to investigate novel ways of interacting with computing systems that improve collaborative data analysis capabilities and decision support assistance in a mission-critical, often time-constrained, work context.

Relevant publications by team members this year: [24], [15], [18], [17], [25].

4.2. Exploratory analysis of scientific data

Many scientific disciplines are increasingly data-driven, including astronomy, molecular biology, particle physics, or neuroanatomy. While making the right decision under time pressure is often less of critical issue when analyzing scientific data, at least not on the same temporal scale as truly time-critical systems, scientists are still faced with large-to-huge amounts of data. No matter their origin (experiments, remote observations, large-scale simulations), these data are difficult to understand and analyze in depth because of their sheer size and complexity. Challenges include how to help scientists freely-yet-efficiently explore their data, keep a trace of the multiple data processing paths they considered to verify their hypotheses and make it easy to backtrack, and how to relate observations made on different parts of the data and insights gained at different moments during the exploration process. With this application area, our goal is to investigate how data-centric interactive systems can improve collaborative scientific data exploration, where users' goals are more open-ended, and where roles, collaboration and coordination patterns [47] differ from those observed in mission-critical contexts of work.

Relevant publications by team members last year: [8].

IMAGINE Project-Team (section vide)

<p style="text-align:center; color:red;">**LACODAM Project-Team**</p>

4. Application Domains

4.1. Introduction

The current period is extremely favorable for teams working in Data Science and Artificial Intelligence, and LACODAM is not the exception. We are eager to see our work applied in real world applications, and have thus an important activity in maintaining strong ties with industrials partners concerned with marketing and energy as well as public partners working on health, agriculture and environment.

4.2. Industry

We present below our industrial collaborations. Some are well established partnerships, while others are more recent collaborations with local industries that wish to reinforce their Data Science R&D with us (e.g. STMicroelectronics, Energiency, Amossys).

- **Execution trace analysis for SOC debugging (STMicroelectronics)**. We have an ongoing collaborations with STMicroelectronics, which is one of top-5 electronic chip makers worldwide. Nowadays, set-top boxes, smartphones and onboard car computers are powered by highly integrated chips called System-on-Chip (SoC). Such chips contain on a single die, processing units, memories, IO units and specialized accelerators (such as audio and video encoding/decoding). Programming SoC is a hard task due to their inherent parallelism, leading to subtle bugs when several components do not deliver their results within a given time frame. Existing debuggers and profilers are ill-adapted in this case because of their high intrusivity that modifies the timings. Hence the most used technique is to capture a trace of the execution and analyze it post-mortem. While Alexandre Termier was in Grenoble he initiated several works for analyzing such traces with pattern mining techniques, which he is now pursuing with his colleagues of the LACODAM project-team.

- **Resource consumption analysis for optimizing energy consumption and practices in industrial factories (Energiency)**. In order to increase their benefits, companies introduce more and more sensors in their factories. Thus, the resource (electricity, water, etc.) consumption of engines, workshops and factories are recorded in the form of times series or temporal sequences. The person who is in charge of resource consumption optimization needs better software than classical spreadsheets for this purpose. He/she needs effective decision-aiding tools with statistical and artificial intelligence knowledge. The start-up Energiency aims at designing and offering such pieces of software for analyzing energy consumption. The starting CIFRE PhD thesis of Maël Guillemé aims at proposing new approaches and solutions from the data mining field to tackle this issue.

- **Security (Amossys)**. Current networks are faced with an increasing variety of attacks, from the classic "DDoS" that makes a server unusable for a few fours, to advanced attacks that silently infiltrate a network and exfiltrate sensitive information months or even years later. Such intrusions, called APT (Advanced Persistent Threat) are extremely hard to detect, and this will become even harder as most communications will be encrypted. A promising solution is to work on "behavioral analysis", by discovering patterns based on the metadata of IP-packets. Such patterns can relate to an unusual sequencing of events, or to an unusual communication graph. Finding such complex patterns over a large volume of streaming data requires to revisit existing stream mining algorithms to dramatically improve their throughput, while guaranteeing a manageable false positive rate. We are collaborating on this topic with the Amossys company and the EMSEC team of Irisa through the co-supervision of a CIFRE PhD (located in the EMSEC team). Our goal is to design novel anomaly detection methods that can detect APT, and that scales on real traffic volumes.

- **Market basket data analysis (Intermarché) and multi-channel interaction data analysis (EDF) for better Customer Relationship Management (CRM)**. An important application domain of data mining for companies that deal with large numbers of customers is to analyze customer interaction data, either for marketing purposes or to improve the quality of service. We have activities in both settings. In the first case, we collaborate with a major french retailer, Intermarché, in order to detect customer churn by analyzing market basket data. In the second case, we collaborate with the major french power supplier, EDF, to discover actionable patterns for CRM that aim at avoiding undesirable situations. We use logs of user interactions with the company (e.g., web clicks, phone calls, etc.) for this purpose.

4.3. Health

- **Care pathways analysis for supporting pharmaco-epidemiological studies**. Pharmaco-epidemiology applies the methodologies developed in general epidemiology to answer to questions about the uses and effects of health products, drugs [33], [32] or medical devices [27], on population. In classical pharmaco-epidemiology studies, people who share common characteristics are recruited to build a dedicated prospective cohort. Then, meaningful data (drug exposures, diseases, etc.) are collected from the cohort within a defined period of time. Finally, a statistical analysis highlights the links (or the lack of links) between drug exposures and outcomes (*e.g.* adverse effects). The main drawback of prospective cohort studies is the time required to collect the data and to integrate them. Indeed, in some cases of health product safety, health authorities have to answer quickly to pharmaco-epidemiology questions.

  New approaches of pharmaco-epidemiology consist in using large EHR (Electronic Health Records) databases to investigate the effects and uses (or misuses) of drugs in real conditions. The objective is to benefit from nationwide available data to answer accurately and in a short time pharmaco-epidemiological queries for national public health institutions. Despite the potential availability of the data, their size and complexity make their analysis long and tremendous. The challenge we tackle is the conception of a generic digital toolbox to support the efficient design of a broad range of pharmaco-epidemiology studies from EHR databases.

  We propose to use pattern mining algorithm and reasoning techniques to analyse the typical care pathways of specific groups of patients.

  To answer the broad range of pharmaco-epidemiological queries from national public health institutions, the PEPS [0] platform exploits, in secondary use, the French health cross-schemes insurance system, called SNDS. The SNDS covers most of the French population with a sliding period of 3 past years. The main characteristics of this data warehouse are described in [31]. Contrary to local hospital EHR or even to other national initiatives, the SNDS data warehouse covers a huge population. It makes possible studies on unfrequent drugs or diseases in real conditions of use. To tackle the volume and the diversity of the SNDS data warehouse, a research program has been established to design an innovative toolbox. This research program is focused first on the modeling of care pathways from the SNDS database and, second, on the design of tools supporting the extraction of insights about massive and complex care pathways by clinicians. In such a database a care pathway is an individual sequence of drugs exposures, medical procedures and hospitalizations.

4.4. Agriculture and environment

- **Dairy farming**. The use and analysis of data acquired in dairy farming is a challenge both for data science and animal science. The goal is to improve farming conditions, i.e., health, welfare and environment, as well as farmers' income. Nowadays, animals are monitored by multiple sensors giving a wealth of heterogeneous data, ex: temperature, weight, or milk composition. Current techniques used by animal scientists focus mostly on mono-sensor approaches. The dynamic

---

[0]PEPS: Pharmaco-Epidémiologie et Produits de Santé – Pharmacoepidemiology of health products

combination of several sensors could provide new services and information useful for dairy farming. The PhD thesis of Kevin Fauvel (#DigitAg grant), aims to study such combinations of sensors and to investigate the use data mining methods, especially pattern mining algorithms. The challenge is to design new algorithms that take into account the data heterogeneity –in terms of nature and time units–, and that produce useful patterns for dairy farming. The outcome of this thesis will be an original and important contribution to the new challenge of the IoT (Internet of Things) and will interest domain actors to find new added value to a global data analysis. The PhD thesis, started on October 2017, takes place in an interdisciplinary setting bringing together computer scientists from Inria and animal scientists from INRA, both located in Rennes.

Similar problems are investigated with the veterinary department of the University of Calgary in the context of cattle monitoring from multiple sensors placed on calves for the early detection of diseases.

- **Optimizing the nutrition of individual sow**. Another direction for further research is the combination of data flows with prediction models in order to learn nutrition strategies. Raphaël Gauthier started a PhD thesis (#DigitAg Grant) in November 2017 with both Inria and INRA supervisors. His research addresses the problem of finding the optimal diet to be supplied to individual sows. Given all the information available, e.g., time-series information about previous feeding, environmental data, scientists models, the research goal is to design new algorithms to determine the optimal ration for a given sow in a given day. Efficiency issues of developed algorithms will be considered since the proposed software should work in real-time on the automated feeder. The decision support process should involve the stakeholder to ensure a good level of acceptance, confidence and understanding of the final tool.

- **Ecosystem modeling and management**. Ongoing research on ecosystem management includes modelling of ecosystems and anthroprogenic pressures, with a special concern on the representation of socio-economical factors that impact human decisions. A main research issue is how to represent these factors and how to integrate their impact on the ecosystem simulation model. This work is an ongoing cooperation with ecologists from the Marine Spatial Ecology of Queensland University, Australia and from Agrocampus Ouest.

LAGADIC Project-Team

4. Application Domains

4.1. Application Domains

The natural applications of our research are obviously in robotics. In fact, researches undertaken in the Lagadic group can apply to all the fields of robotics implying a vision sensor. They are indeed conceived to be independent of the system considered (and the robot and the vision sensor can even be virtual for some applications).

Currently, we are mostly interested in using visual servoing for aerial and space application, micromanipulation, autonomous vehicle navigation in large urban environments or for disabled or elderly people.

We also address the field of medical robotics. The applications we consider turn around new functionalities of assistance to the clinician during a medical examination: visual servoing on echographic images, needle insertion, compensation of organ motion, etc.

Robotics is not the only possible application field to our researches. In the past, we were interested in applying visual servoing in computer animation, either for controlling the motions of virtual humanoids according to their pseudo-perception, or for controlling the point of view of visual restitution of an animation. In both cases, potential applications are in the field of virtual reality, for example for the design of video games, or virtual cinematography.

Applications also exist in computer vision and augmented reality. It is then a question of carrying out a virtual visual servoing for the 3D localization of a tool with respect to the vision sensor, or for the estimation of its 3D motion. This field of application is very promising, because it is in full rise for the realization of special effects in the multi-media field or for the design and the inspection of objects manufactured in the industrial world.

LARSEN Project-Team

4. Application Domains

4.1. Personal Assistance

During the last fifty years, many medical advances as well as the improvement of the quality of life have resulted in a longer life expectancy in industrial societies. The increase in the number of elderly people is a matter of public health because although elderly people can age in good health, old age also causes embrittlement, in particular on the physical plan which can result in a loss of autonomy. That will force us to re-think the current model regarding the care of elderly people. [0] Capacity limits in specialized institutes, along with the preference of elderly people to stay at home as long as possible, explain a growing need for specific services at home.

Ambient intelligence technologies and robotics could contribute to this societal challenge. The spectrum of possible actions in the field of elderly assistance is very large. We will focus on activity monitoring services, mobility or daily activity aids, medical rehabilitation, and social interactions. This will be based on the experimental infrastructure we have build in Nancy (Smart apartment platform) as well as the deep collaboration we have with OHS. [0]

4.2. Civil Robotics

Many applications for robotics technology exist within the services provided by national and local government. Typical applications include civil infrastructure services [0] such as: urban maintenance and cleaning; civil security services; emergency services involved in disaster management including search and rescue; environmental services such as surveillance of rivers, air quality, and pollution. These applications may be carried out by a wide variety of robot and operating modality, ranging from single robots or small fleets of homogeneous or heterogeneous robots. Often robot teams will need to cooperate to span a large workspace, for example in urban rubbish collection, and operate in potentially hostile environments, for example in disaster management. These systems are also likely to have extensive interaction with people and their environments.

The skills required for civil robots match those developed in the LARSEN project: operating for a long time in potentially hostile environment, potentially with small fleets of robots, and potentially in interaction with people.

---

[0]See the Robotics 2020 Multi-Annual Roadmap [57].
[0]OHS (*Office d'Hygiène Sociale*) is an association managing several rehabilitation or retirement home structures.
[0]See the Robotics 2020 Multi-Annual Roadmap [57], section 2.5.

<p style="text-align: center; color: red;">**LINKMEDIA Project-Team**</p>

4. Application Domains

4.1. Asset management in the entertainement business

Regardless of the ingestion and storage issues, media asset management—archiving, describing and retrieving multimedia content—has turned into a key factor and a huge business for content and service providers. Most content providers, with television channels at the forefront, rely on multimedia asset management systems to annotate, describe, archive and search for content. So do archivists such as the Institut National de l'Audiovisuel, the Nederlands Instituut voor Beeld en Geluid or the British Broadcast Corporation, as well as media monitoring companies, such as Yacast in France. Protecting copyrighted content is another aspect of media asset management.

4.2. Multimedia Internet

One of the most visible application domains of linked multimedia content is that of multimedia portals on the Internet. Search engines now offer many features for image and video search. Video sharing sites also feature search engines as well as recommendation capabilities. All news sites provide multimedia content with links between related items. News sites also implement content aggregation, enriching proprietary content with user-generated content and reactions from social networks. Most public search engines and Internet service providers offer news aggregation portals.

4.3. Multiscreen TV

The convergence between television and the Internet has accelerated significantly over the past few years, with the democratization of TV on-demand and replay services and the emergence of social TV services and multiscreen applications. These evolutions and the consequently ever growing number of innovative applications offer a unique playground for multimedia technologies. Recommendation plays a major role in connected TV. Enriching multimedia content, with explicit links targeting either multimedia material or knowledge databases, appears as a key feature in this context, at the core of rich TV and second screen applications.

4.4. E-learning

On-line courses are rapidly gaining interest with the recent movement for massive open on-line courses (MOOCs). Such courses usually aggregate multimedia material, such as a video of the course with handouts and potentially text books, exercises and other related resources. This setting is very similar to that of the media aggregation sites though in a different domain. Automatically analyzing and describing video and textual content, synchronizing all material available across modalities, creating and characterizing links between related material or between different courses are all necessary features for on-line courses authoring.

<p style="text-align:center; color:red;">**LINKS Project-Team**</p>

4. Application Domains

4.1. Linked Data Integration

There are many contexts in which integrating linked data is interesting. We advocate here one possible scenario, namely that of integrating business linked data to feed what is called Business Intelligence. The latter consists of a set of theories and methodologies that transform raw data into meaningful and useful information for business purposes (from Wikipedia). In the past decade, most of the enterprise data was proprietary, thus residing within the enterprise repository, along with the knowledge derived from that data. Today's' enterprises and businessmen need to face the problem of information explosion, due to the Internet's ability to rapidly convey large amounts of information throughout the world via end-user applications and tools. Although linked data collections exist by bridging the gap between enterprise data and external resources, they are not sufficient to support the various tasks of Business Intelligence. To make a concrete example, concepts in an enterprise repository need to be matched with concepts in Wikipedia and this can be done via pointers or equalities. However, more complex logical statements (i.e. mappings) need to be conceived to map a portion of a local database to a portion of an RDF graph, such as a subgraph in Wikipedia or in a social network, e.g. LinkedIn. Such mappings would then enrich the amount of knowledge shared within the enterprise and let more complex queries be evaluated. As an example, businessmen with the aid of business intelligence tools need to make complex sentimental analysis on the potential clients and for such a reason, such tools must be able to pose complex queries, that exploit the previous logical mappings to guide their analysis. Moreover, the external resources may be rapidly evolving thus leading to revisit the current state of business intelligence within the enterprise.

4.2. Data Cleaning

The second example of application of our proposal concerns scientists who want to quickly inspect relevant literature and datasets. In such a case, local knowledge that comes from a local repository of publications belonging to a research institute (e.g. HAL) need to be integrated with other Web-based repositories, such as DBLP, Google Scholar, ResearchGate and even Wikipedia. Indeed, the local repository may be incomplete or contain semantic ambiguities, such as mistaken or missing conference venues, mistaken long names for the publication venues and journals, missing explanation of research keywords, and opaque keywords. We envision a publication management system that exploits both links between database elements, namely pointers to external resources and logical links. The latter can be complex relationships between local portions of data and remote resources, encoded as schema mappings. There are different tasks that such a scenario could entail such as (i) cleaning the errors with links to correct data e.g. via mappings from HAL to DBLP for the publications errors, and via mappings from HAL to Wikipedia for opaque keywords, (ii) thoroughly enrich the list of publications of a given research institute, and (iii) support complex queries on the corrected data combined with logical mappings.

4.3. Real Time Complex Event Processing

Complex event processing serves for monitoring nested word streams in real time. Complex event streams are gaining popularity with social networks such as with Facebook and Twitter, and thus should be supported by distributed databases on the Web. Since this is not yet the case, there remains much space for future industrial transfer related to Links' second axis on dynamic linked data.

MAGNET Project-Team

4. Application Domains

4.1. Domain

Our main targeted applications are browsing, monitoring, recommending and mining in information networks. The learning tasks considered in the project such as node clustering, node and link classification and link prediction are likely to yield important improvements in these applications. Application domains cover social networks for cultural data and e-commerce, and biomedical informatics.

We also target applications related to decentralized learning and privacy preserving systems when users or devices are interconnected in large networks. We develop solutions based on urban and mobility data where privacy is a specific requirement.

4. Application Domains

4.1. Augmented reality

We have a significant experience in AR that allowed good progress in building usable, reliable and robust AR systems. Our contributions cover the entire process of AR: matching, pose initialization, 3D tracking, in-situ modeling, handling interaction between real and virtual objects....

4.2. Medical Imaging

For 15 years, we have been working in close collaboration with the University Hospital of Nancy and GE Healthcare in interventional neuroradiology. Our common aim is to develop a multimodal framework to help therapeutic decisions and interventional gestures. Contributions of the team focus on the developments of AR tools for neuro-navigation as well as the development of simulation tools for training or planning. Laparoscopic surgery is another field of interest with the development of methods for tracking deformable organs based on bio-mechanical models. Some of these projects are developed in collaboration with the MIMESIS project team.

4.3. Experimental mechanics

In experimental solid mechanics, an important problem is to characterize properties of specimen subject to mechanical constraints, which makes it necessary to measure tiny strains. Contactless measurement techniques have emerged in the last few years and are spreading quickly. They are mainly based on images of the surface of the specimen on which a regular grid or a random speckle has been deposited. We are engaged since June 2012 in a transdisciplinary collaboration with Institut Pascal (Clermont Auvergne Université). The aim is to characterize the metrological performances of these techniques limited by, e.g., the sensor noise, and to improve them by several dedicated image processing tools.

MANAO Project-Team

4. Application Domains

4.1. Physical Systems

Given our close relationships with researchers in optics, one novelty of our approach is to extend the range of possible observers to physical sensors in order to work on domains such as simulation, mixed reality, and testing. Capturing, processing, and visualizing complex data is now more and more accessible to everyone, leading to the possible convergence of real and virtual worlds through visual signals. This signal is traditionally captured by cameras. It is now possible to augment them by projecting (e.g., the infrared laser of Microsoft Kinect) and capturing (e.g., GPS localization) other signals that are outside the visible range. This supplemental information replaces values traditionally extracted from standard images and thus lowers down requirements in computational power. Since the captured images are the result of the interactions between light, shape, and matter, the approaches and the improved knowledge from *MANAO* help in designing interactive acquisition and rendering technologies that are required to merge the real and the virtual worlds. With the resulting unified systems (optical and digital), transfer of pertinent information is favored and inefficient conversion is likely avoided, leading to new uses in interactive computer graphics applications, like **augmented reality**, **displays** and **computational photography**.

4.2. Interactive Visualization and Modeling

This direction includes domains such as **scientific illustration and visualization**, **artistic or plausible rendering**, and **3D modeling**. In all these cases, the observer, a human, takes part in the process, justifying once more our focus on real-time methods. When targeting average users, characteristics as well as limitations of the human visual system should be taken into account: in particular, it is known that some configurations of light, shape, and matter have masking and facilitation effects on visual perception. For specialized applications (such as archeology), the expertise of the final user and the constraints for 3D user interfaces lead to new uses and dedicated solutions for models and algorithms.

MAVERICK Project-Team

4. Application Domains

4.1. Application Domains

The natural application domain for our research is the production of digital images, for example for movies and special effects, virtual prototyping, video games...

Our research have also been applied to tools for generating and editing images and textures, for example generating textures for maps.

Our current application domains are:

- Offline and real-time rendering in movie special effects and video games;
- Virtual prototyping;
- Scientific visualization;
- Content modeling and generation (e.g. generating texture for video games, capturing reflectance properties, etc);
- Image creation and manipulation.

MIMETIC Project-Team

4. Application Domains

4.1. Autonomous Characters

Autonomous characters are becoming more and more popular as they are used in an increasing number of application domains. In the field of special effects, virtual characters are used to replace secondary actors and generate highly populated scenes that would be hard and costly to produce with real actors. In video games and virtual storytelling, autonomous characters play the role of actors that are driven by a scenario. Their autonomy allows them to react to unpredictable user interactions and adapt their behavior accordingly. In the field of simulation, autonomous characters are used to simulate the behavior of humans in different kind of situations. They enable to study new situations and their possible outcomes.

One of the main challenges in the field of autonomous characters is to provide a unified architecture for the modeling of their behavior. This architecture includes perception, action and decisional parts. This decisional part needs to mix different kinds of models, acting at different time scale and working with different nature of data, ranging from numerical (motion control, reactive behaviors) to symbolic (goal oriented behaviors, reasoning about actions and changes).

In the MimeTIC team, we focus on autonomous virtual humans. Our problem is not to reproduce the human intelligence but to propose an architecture making it possible to model credible behaviors of anthropomorphic virtual actors evolving/moving in real time in virtual worlds. The latter can represent particular situations studied by psychologists of the behavior or to correspond to an imaginary universe described by a scenario writer. The proposed architecture should mimic all the human intellectual and physical functions.

4.2. Biomechanics and Motion Analysis

Biomechanics is obviously a very large domain. This large set can be divided regarding to the scale at which the analysis is performed going from microscopic evaluation of biological tissues' mechanical properties to macroscopic analysis and modeling of whole body motion. Our topics in the domain of biomechanics mainly lie within this last scope. In order to obtain a better understanding of human motion, MimeTIC addresses three main situations: everyday motions of a lambda subject, locomotion of pathological subjects and sports gestures.

In the first situation, MimeTIC is interested in studying how subjects maintain their balance in highly dynamic conditions. Until now, balance have nearly always been considered in static or quasi-static conditions. The knowledge of much more dynamic cases still has to be improved. Our approach has demonstrated that, first of all, the question of the parameter that will allow to do this is still open. We have also largely contributed to gaining a better understanding of collision avoidance between pedestrians. This topic includes the research of the parameters that are interactively controlled and the study of each one's role within this interaction.

The second situation focuses on locomotion of pathological subjects. When patients cannot walk efficiently, in particular those suffering from central nervous system affections, it becomes very useful for practitioners to benefit from an objective evaluation of their capacities. To facilitate such evaluations, we have developed two complementary indices, one based on kinematics and the other one on muscle activations. One major point of our research is that such indices are usually only developed for children whereas adults with these affections are much more numerous. Finally, in sports, where gesture can be considered, in some way, as abnormal, the goal is more precisely to understand the determinants of performance. This could then be used to improve training programs or devices. Two different sports have been studied: a) the tennis serve, where the goal was to understand the contribution of each segment of the body on the speed of the ball and b) the influence of the mechanical characteristics of the fin in fin swimming.

After having improved the knowledge of these different gestures a second goal is then to propose modeling solutions that can be used in VR environments for other research topics within MimeTIC. This has been the case, for example, for collision avoidance.

4.3. Interactions between walkers

Modeling and simulating the interactions between walkers is a very active, complex and competitive domain, interesting various disciplines such as mathematics, cognitive sciences, physics, computer graphics, etc. Interactions between walkers are by definition at the very core of our society since they represent the basic synergies of our daily life. When walking in the street, we take information about our surrounding environment in order to interact with people, move without collision, alone or in a group, intercept, meet or escape to somebody. Large groups of walkers can be first seen as a complex system: numerous local interactions occur between its elements and result into macroscopic emergent phenomena. Interactions are of various nature (e.g., collision avoidance, following) and are undergoing various factors as well. Physical factors are crucial as a group gathers by definition numerous moving people with a certain level of density. But sociological, cultural and psychological factors are important as well, since people's behavior is deeply changed from country to country, or depending on the considered situations. On the computational point of view, simulating the movements of large groups of walkers (i.e., crowds) pushes traditional simulation algorithms to their limit. As an element of a crowd is subject to interact with any other element belonging the same crowd, a naïve simulation algorithm has a quadratic complexity. Specific strategies are set to face such a difficulty: level-of-detail techniques enable scaling large crowd simulation and reach real-time solutions.

MimeTIC is an international key contributor in the domain of understanding and simulating interactions between walkers, in particular for virtual crowds. Our approach is specific and based on three axes. First, our modeling approach is based on human movement science: we conduct challenging experiments focusing on the perception as well as on the motion involved in local interactions between walkers both using real and virtual set-ups. Second: we develop high-performance solutions for crowd simulation. Third, we develop solutions for realistic navigation in virtual world to enable interaction with crowds in Virtual Reality.

4.4. Motion Sensing of Human Activity

Recording human activity is a key point of many applications and fundamental works. Numerous sensors and systems have been proposed to measure positions, angles or accelerations of the user's body parts. Whatever the system is, one of the main problems is to be able to automatically recognize and analyze the user's performance according to poor and noisy signals. Human activity and motion are subject to variability: intra-variability due to space and time variations of a given motion, but also inter-variability due to different styles and anthropometric dimensions. MimeTIC has addressed the above problems in two main directions.

Firstly, we have studied how to recognize and quantify motions performed by a user when using accurate systems such as Vicon (product of Oxford Metrics) or Optitrack (product of Natural Point) motion capture systems. These systems provide large vectors of accurate information. Due to the size of the state vector (all the degrees of freedom) the challenge is to find the compact information (named features) that enables the automatic system to recognize the performance of the user. Whatever the method used, finding these relevant features that are not sensitive to intra-individual and inter-individual variability is a challenge. Some researchers have proposed to manually edit these features (such as a Boolean value stating if the arm is moving forward or backward) so that the expertise of the designer is directly linked with the success ratio. Many proposals for generic features have been proposed, such as using Laban notation which was introduced to encode dancing motions. Other approaches tend to use machine learning to automatically extract these features. However most of the proposed approaches were used to seek a database for motions which properties correspond to the features of the user's performance (named motion retrieval approaches). This does not ensure the retrieval of the exact performance of the user but a set of motions with similar properties.

Secondly, we wish to find alternatives to the above approach which is based on analyzing accurate and complete knowledge on joint angles and positions. Hence new sensors, such as depth-cameras (Kinect, product of Microsoft) provide us with very noisy joint information but also with the surface of the user. Classical approaches would try to fit a skeleton into the surface in order to compute joint angles which, again, lead to large state vectors. An alternative would be to extract relevant information directly from the raw data, such as the surface provided by depth cameras. The key problem is that the nature of these data may be very different from classical representation of human performance. In MimeTIC, we try to address this problem in specific application domains that require picking specific information, such as gait asymmetry or regularity for clinical analysis of human walking.

4.5. VR and Sports

Sport is characterized by complex displacements and motions. These motions are dependent on visual information that the athlete can pick up in his environment, including the opponent's actions. Perception is thus fundamental to the performance. Indeed, a sportive action, as unique, complex and often limited in time, requires a selective gathering of information. This perception is often seen as a prerogative for action, it then takes the role of a passive collector of information. However, as mentioned by Gibson in 1979, the perception-action relationship should not be considered sequential but rather as a coupling: we perceive to act but we must act to perceive. There would thus be laws of coupling between the informational variables available in the environment and the motor responses of a subject. In other words, athletes have the ability to directly perceive the opportunities of action directly from the environment. Whichever school of thought considered, VR offers new perspectives to address these concepts by complementary using real time motion capture of the immersed athlete.

In addition to better understanding sports and interactions between athletes, VR can also be used as a training environment as it can provide complementary tools to coaches. It is indeed possible to add visual or auditory information to better train an athlete. The knowledge found in perceptual experiments can be for example used to highlight the body parts that are important to look at to correctly anticipate the opponent's action.

4.6. Interactive Digital Storytelling

Interactive digital storytelling, including novel forms of edutainment and serious games, provides access to social and human themes through stories which can take various forms and contains opportunities for massively enhancing the possibilities of interactive entertainment, computer games and digital applications. It provides chances for redefining the experience of narrative through interactive simulations of computer-generated story worlds and opens many challenging questions at the overlap between computational narratives, autonomous behaviours, interactive control, content generation and authoring tools.

Of particular interest for the MimeTIC research team, virtual storytelling triggers challenging opportunities in providing effective models for enforcing autonomous behaviours for characters in complex 3D environments. Offering both low-level capacities to characters such as perceiving the environments, interacting with the environment and reacting to changes in the topology, on which to build higher-levels such as modelling abstract representations for efficient reasoning, planning paths and activities, modelling cognitive states and behaviours requires the provision of expressive, multi-level and efficient computational models. Furthermore virtual storytelling requires the seamless control of the balance between the autonomy of characters and the unfolding of the story through the narrative discourse. Virtual storytelling also raises challenging questions on the conveyance of a narrative through interactive or automated control of the cinematography (how to stage the characters, the lights and the cameras). For example, estimating visibility of key subjects, or performing motion planning for cameras and lights are central issues for which have not received satisfactory answers in the literature.

4.7. VR and Ergonomics

The design of workstations nowadays tends to include assessment steps in a Virtual Environment (VE) to evaluate ergonomic features. This approach is more cost-effective and convenient since working directly on the Digital Mock-Up (DMU) in a VE is preferable to constructing a real physical mock-up in a Real Environment (RE). This is substantiated by the fact that a Virtual Reality (VR) set-up can be easily modified, enabling quick adjustments of the workstation design. Indeed, the aim of integrating ergonomics evaluation tools in VEs is to facilitate the design process, enhance the design efficiency, and reduce the costs.

The development of such platforms asks for several improvements in the field of motion analysis and VR. First, interactions have to be as natural as possible to properly mimic the motions performed in real environments. Second, the fidelity of the simulator also needs to be correctly evaluated. Finally, motion analysis tools have to be able to provide in real-time biomechanics quantities usable by ergonomists to analyse and improve the working conditions.

MINT2 Team

4. Application Domains

4.1. Domain

MINT team currently devotes application to domains such as cultural heritage, retail, and health.

<p style="text-align:center; color:red">Mjolnir Team</p>

4. Application Domains

4.1. Application Domains

Mjolnir works on fundamental aspects of Human-Computer Interaction that can be applied to diverse application domains. Our 2017 research concerned desktop, touch-based, haptics, and BCI interfaces with notable applications to education, genetics research, as well as creativity support tools (3D audio production and design of Digital Musical Instruments).

MOEX Project-Team (section vide)

<div align="center">

MORPHEO Project-Team

</div>

4. Application Domains

4.1. 4D modeling

Modeling shapes that evolve over time, analyzing and interpreting their motion has been a subject of increasing interest of many research communities including the computer vision, the computer graphics and the medical imaging communities. Recent evolutions in acquisition technologies including 3D depth cameras (Time-of-Flight and Kinect), multi-camera systems, marker based motion capture systems, ultrasound and CT scans have made those communities consider capturing the real scene and their dynamics, create 4D spatio-temporal models, analyze and interpret them. A number of applications including dense motion capture, dynamic shape modeling and animation, temporally consistent 3D reconstruction, motion analyzes and interpretation have therefore emerged.

4.2. Shape Analysis

Most existing shape analysis tools are local, in the sense that they give local insight about an object's geometry or purpose. The use of both geometry and motion cues makes it possible to recover more global information, in order to get extensive knowledge about a shape. For instance, motion can help to decompose a 3D model of a character into semantically significant parts, such as legs, arms, torso and head. Possible applications of such high-level shape understanding include accurate feature computation, comparison between models to detect defects or medical pathologies, and the design of new biometric models.

4.3. Human Motion Analysis

The recovery of dense motion information enables the combined analyses of shapes and their motions. Typical examples include the estimation of mean shapes given a set of 3D models or the identification of abnormal deformations of a shape given its typical evolutions. The interest arises in several application domains where temporal surface deformations need to be captured and analysed. It includes human body analyses for which potential applications are anyway numerous and important, from the identification of pathologies to the design of new prostheses.

4.4. Interaction

The ability to build models of humans in real time allows to develop interactive applications where users interact with virtual worlds. For instance, Kinect proposed by Microsoft illustrates this principle with game applications using human inputs perceived with a depth camera. Other examples include gesture interfaces using visual inputs. A challenging issue in this domain is the ability to capture complex scenes in natural environments. Multi-modal visual perception, e.g. depth and color cameras, is one objective in that respect.

<p style="text-align:center">MULTISPEECH Project-Team</p>

4. Application Domains

4.1. Introduction

Approaches and models developed in the MULTISPEECH project are intended to be used for facilitating oral communication in various situations through enhancements of the communication channels, either directly via automatic speech recognition or speech production technologies, or indirectly, thanks to computer assisted language learning. Applications also include the usage of speech technologies for helping people in handicapped situations or for improving their autonomy. Foreseen application domains are related to computer assisted learning, health and autonomy (more precisely aided communication and monitoring), annotation and processing of spoken documents, and multimodal computer interaction.

4.2. Computer Assisted Learning

Although speaking seems quite natural, learning foreign languages, or learning the mother tongue for people with language deficiencies, represents critical cognitive stages. Hence, many scientific activities have been devoted to these issues either from a production or a perception point of view. The general guiding principle with respect to computer assisted mother or foreign language learning is to combine modalities or to augment speech to make learning easier. Based upon a comparison of the learner's production to a reference, automatic diagnoses of the learner's production can be considered, as well as perceptual feedback relying on an automatic transformation of the learner's voice. The diagnosis step strongly relies on the studies on categorization of sounds and prosody in the mother tongue and in the second language. Furthermore, reliable diagnosis on each individual utterance is still a challenge, and elaboration of advanced automatic feedback requires a temporally accurate segmentation of speech utterances into phones and this explains why accurate segmentation of native and non-native speech is an important topic in the field of acoustic speech modeling.

4.3. Aided Communication and Monitoring

A foreseen application aims at improving the autonomy of elderly or disabled people, and fit with smartroom applications. In a first step, source separation techniques could be tuned and should help for locating and monitoring people through the detection of sound events inside apartments. In a longer perspective, adapting speech recognition technologies to the voice of elderly people should also be useful for such applications, but this requires the recording of adequate databases. Sound monitoring in other application fields (security, environmental monitoring) could also be envisaged.

4.4. Annotation and Processing of Spoken Documents and Audio Archives

A first type of annotation consists in transcribing a spoken document in order to get the corresponding sequences of words, with possibly some complementary information, such as the structure (punctuation) or the modality (affirmation/question) of the utterances to make the reading and understanding easier. Typical applications of the automatic transcription of radio or TV shows, or of any other spoken document, include making possible their access by deaf people, as well as by text-based indexing tools.

A second type of annotation is related to speech-text alignment, which aims at determining the starting and ending times of the words, and possibly of the sounds (phonemes). This is of interest in several cases as for example, for annotating speech corpora for linguistic studies, and for synchronizing lip movements with speech sounds, for example for avatar-based communications. Although good results are currently achieved on clean data, automatic speech-text alignment needs to be improved for properly processing noisy spontaneous speech data and needs to be extended to handle overlapping speech.

Large audio archives are important for some communities of users, e.g., linguists, ethnologists or researchers in digital humanities in general. In France, a notorious example is the "Archives du CNRS — Musée de l'homme", gathering about 50,000 recordings dating back to the early 1900s. When dealing with very old recordings, the practitioner is often faced with the problem of noise. This stems from the fact that a lot of interesting material from a scientific point of view is very old or has been recorded in very adverse noisy conditions, so that the resulting audio is poor. The work on source separation can lead to the design of semi-automatic denoising and enhancement features, that would allow these researchers to significantly enhance their investigation capabilities, even without expert knowledge in sound engineering.

Finally, there is also a need for speech signal processing techniques in the field of multimedia content creation and rendering. Relevant techniques include speech and music separation, speech equalization, prosody modification, and speaker conversion.

4.5. Multimodal Computer Interactions

Speech synthesis has tremendous applications in facilitating communication in a human-machine interaction context to make machines more accessible. For example, it started to be widely common to use acoustic speech synthesis in smartphones to make possible the uttering of all the information. This is valuable in particular in the case of handicap, as for blind people. Audiovisual speech synthesis, when used in an application such as a talking head, i.e., virtual 3D animated face synchronized with acoustic speech, is beneficial in particular for hard-of-hearing individuals. This requires an audiovisual synthesis that is intelligible, both acoustically and visually. A talking head could be an intermediate between two persons communicating remotely when their video information is not available, and can also be used in language learning applications as vocabulary tutoring or pronunciation training tool. Expressive acoustic synthesis is of interest for the reading of a story, such as audiobook, to facilitate the access to literature (for instance for blind people or illiterate people).

<p align="center" style="color:red">ORPAILLEUR Project-Team</p>

4. Application Domains

4.1. Life Sciences: Biology, Chemistry and Medicine

Participants: Adrien Coulet, Kévin Dalleau, Esther Catherine Galbrun, Nicolas Jay, Joël Legrand, Jean Lieber, Pierre Monnin, Amedeo Napoli, Chedy Raïssi, Mohsen Sayed, Malika Smaïl-Tabbone, Yannick Toussaint.

> **Keywords:** knowledge discovery in life sciences, biology, chemistry, medicine, pharmacogenomics and precision medicine.

One major application domain which is currently investigated by the Orpailleur team is related to life sciences, with particular emphasis on biology, medicine, and chemistry. The understanding of biological systems provides complex problems for computer scientists, and the developed solutions bring new research ideas or possibilities for biologists and for computer scientists as well. Indeed, the interactions between researchers in biology and researchers in computer science improve not only knowledge about systems in biology, chemistry, and medicine, but knowledge about computer science as well.

Knowledge discovery is gaining more and more interest and importance in life sciences for mining either homogeneous databases such as protein sequences and structures, or heterogeneous databases for discovering interactions between genes and environment, or between genetic and phenotypic data, especially for public health and precision medicine (pharmacogenomics). Pharmacogenomics is one main challenge for the Orpailleur team as it considers a large panel of complex data ranging from biological to medical data, and various kinds of encoded domain knowledge ranging from texts to formal ontologies.

On the same line as biological data, chemical data are presenting important challenges w.r.t. knowledge discovery, for example for mining collections of molecular structures and collections of chemical reactions in organic chemistry. The mining of such collections is an important task for various reasons among which the challenge of graph mining and the industrial needs (especially in drug design, pharmacology and toxicology). Molecules and chemical reactions are complex data that can be modeled as labeled graphs. Graph mining methods may play an important role in this framework and Formal Concept Analysis can also be used in an efficient and well-founded way [86]. Graph mining as considered in the framework of FCA is an important task on which we are working, whose results can be transferred to text mining as well.

We are working on knowledge management in medicine and analysis of patient trajectories. The Kasimir research project is about decision support and knowledge management for the treatment of cancer. This is a multidisciplinary research project in which researchers in computer science (Orpailleur) and experts in oncology are participating. For a given cancer localization, a treatment is based on a protocol, which is applied in 70% of the cases and provides a treatment. The 30% remaining cases are "out of the protocol", e.g. contraindication, treatment impossibility, etc. and the protocol have to be adapted, based on discussions among specialists. This adaptation process is modeled in Kasimir thanks to CBR, where semantic web technologies have been used.

The analysis of patient trajectories, i.e. the "path" of a patient during illness (chronic illnesses and cancer), can be considered as an analysis of sequences. It is important to understand such sequential data and sequence mining methods should be adapted for addressing the complex nature of medical events. We are interested in the analysis of trajectories at different levels of granularity and w.r.t. external domain ontologies. In addition, it is also important to be able to compare and classify trajectories according to their content. Then we are also interested in the definition of similarity measures able to take into account the complex nature of trajectories and that can be efficiently implemented for allowing quick and reliable classifications.

4.2. Other Application Domains

Participants: Emmanuelle Gaillard, Florence Le Ber, Jean Lieber, Jean-François Mari, Amedeo Napoli, Emmanuel Nauer, Sébastien Da Silva.

4.2.1. Cooking

Keywords: cooking, knowledge engineering, case-based reasoning, semantic web

The origin of the Taaable project is the Computer Cooking Contest (CCC). A contestant to CCC is a system that answers queries about recipes, using a recipe base; if no recipe exactly matches the query, then the system adapts another recipe. Taaable is a case-based reasoning system based on knowledge representation, semantic web and knowledge discovery technologies. The system enables to validate scientific results and to study the complementarity of various research trends in an application domain which is simple to understand and which raises complex issues at the same time.

4.2.2. Agronomy

Keywords: simulation in agronomy, graph model in agronomy

Research in agronomy was conducted in cooperation between Inria and INRA, within the INRA research network PAYOTE about landscape modeling. The research work was related to the characterization and the simulation of hedgerow structures in agricultural landscapes, based on Hilbert-Peano curves and Markov models [73]. Moreover, an on-going research work about the representation of farmer experience is carried on within a collaboration with IRD in Madagascar [81]. Sketches drawn by farmers were transformed into graphs and compared thanks to Formal Concept Analysis.

4.2.3. Digital Humanities

Keywords: digital humanities, semantic web, SPARQL, approximate search, case-based reasoning

Members of the Orpailleur team are collaborating with a group of researchers working in history and philosophy of science and technologies (they are located in Brest, Montpellier and Nancy). The idea is to reuse semantic web technologies for better access and better representation of their text corpora.

PANAMA Project-Team

4. Application Domains

4.1. Acoustic Scene Capture

Acoustic fields carry much information about audio sources (musical instruments, speakers, etc.) and their environment (e.g., church acoustics differ much from office room acoustics). A particular challenge is to capture as much information from a complete 3D+t acoustic field associated with an audio scene, using as few sensors as possible. The feasibility of compressive sensing to address this challenge was shown in certain scenarii, and the actual implementation of this framework will potentially impact practical scenarii such as remote surveillance to detect abnormal events, e.g. for health care of the elderly or public transport surveillance.

4.2. Audio Signal Separation in Reverberant Environments

Audio signal separation consists in extracting the individual sound of different instruments or speakers that were mixed on a recording. It is now successfully addressed in the academic setting of linear instantaneous mixtures. Yet, real-life recordings, generally associated to reverberant environments, remain an unsolved difficult challenge, especially with many sources and few audio channels. Much of the difficulty comes from the estimation of the unknown room impulse response associated to a matrix of mixing filters, which can be expressed as a dictionary-learning problem. Solutions to this problem have the potential to impact, for example, the music and game industry, through the development of new digital re-mastering techniques and virtual reality tools, but also surveillance and monitoring applications, where localizing audio sources is important.

4.3. Multimedia Indexing

Audiovisual and multimedia content generate large data streams (audio, video, associated data such as text, etc.). Manipulating large databases of such content requires efficient techniques to: segment the streams into coherent sequences; label them according to words, language, speaker identity, and more generally to the type of content; index them for easy querying and retrieval, etc. As the next generation of online search engines will need to offer content-based means of searching, the need to drastically reduce the computational burden of these tasks is becoming all the more important as we can envision the end of the era of wasteful datacenters that can increase forever their energy consumption. Most of today's techniques to deal with such large audio streams involve extracting features such as Mel Frequency Cepstral Coefficients (MFCC) and learning high-dimensional statistical models such as Gaussian Mixture Models, with several thousand parameters. The exploration of a compressive learning framework is expected to contribute to new techniques to efficiently process such streams and perform segmentation, classification, etc., in the compressed domain. A particular challenge is to understand how this paradigm can help exploiting truly multimedia features, which combine information from different associated streams such as audio and video, for joint audiovisual processing.

PERCEPTION Project-Team (section vide)

<p style="text-align:center; color:red;">**PERVASIVE INTERACTION Project-Team**</p>

4. Application Domains

4.1. Smart Energy Systems

Participants: Amr Alyafi, Amine Awada, Patrick Reignier Partners: UMR G-SCOP, UMR LIG (Persuasive Interaction, IIHM), CEA Liten, PACTE, Vesta Systems and Elithis.

Work in this area explores techniques for a user centric energy management system, where user needs and tacit knowledge drive the search of solutions. These are calculated using a flexible energy model of the living areas. The system is personified by energy consultants with which building actors such as building owners, building managers, technical operators but also occupants, can interact in order to co-define energy strategies, benefiting of both assets: tacit knowledge of human actors, and measurement with computation capabilities of calculators. Putting actors in the loop, i.e. making energy not only visible but also controllable is the needed step before large deployment of energy management solutions. It is proposed to develop interactive energy consultants for all the actors, which are energy management aided systems embedding models in order to support the decision making processes. MIRROR (interactive monitoring), WHAT-IF (interactive quantitative simulation), EXPLAIN (interactive qualitative simulation), SUGGEST-AND-ADJUST (interactive management) and RECOMMEND (interactive diagnosis) functionalities will be developed.

4.2. E-Textile

Participant: Sabine Coquillart

Partner: LIMSI

Collaboration with the HAPCO team from LIMSI on e-textiles. A patent application has been filed related to this work:

- F. Bimbard, M. Bobin, M. Ammi, S. Coquillart "Procédé de conception d'un capteur de flexion textile piézorésistif à partir de fils fonctionnels", Patent Application, 2017.

4.3. Interaction with Pervasive Media

Participants: Sabine Coquillart, Jingtao Chen

Partners: Inria GRA, GIPSA, G-SCOP

Pseudo-haptic feedback is a technique aiming to simulate haptic sensations without active haptic feedback devices. Peudo-haptic techniques have been used to simulate various haptic feedbacks such as stiffness, torques, and mass. In the framework of Jingtao Chen PhD thesis, a novel pseudo-haptic experiment has been set up. The aim of this experiment is to study the EMG signals during a pseudo-haptic task. A stiffness discrimination task similar to the one published in Lecuyer's PhD thesis has been chosen. The experimental set-up has been developed, as well as the software controlling the experiment. Pre-tests are under way. They will be followed by the tests with subjects.

4.4. Bayesian Reasoning

Participants: Emmanuel Mazer, Marvin Faix

The development of modern computers is mainly based on increase of performances and decrease of size and energy consumption, with no notable modification of the basic principles of computation. In particular, all the components perform deterministic and exact operations on sets of binary signals. These constraints obviously impede further sizable progresses in terms of speed, miniaturization and power consumption. The main goal of the project MicroBayes is to investigate a radically different approach, using stochastic bit streams to perform computations. The aim of this project is to show that stochastic architectures can outperform standard computers to solve complex inference problems both in terms of execution speed and of power consumption. We will demonstrate the feasibility on two applications involving low level information processing from sensor signals, namely sound source localization and separation.

<p style="text-align:center"><strong style="color:red">PETRUS Project-Team</p>

4. Application Domains

4.1. Application Domains

As stated in the software section, the Petrus research strategy aims at materializing its scientific contributions in an advanced hardware/software platform with the expectation to produce a real societal impact. Hence, our software activity is structured around a common Secure Personal Cloud platform rather than several isolated demonstrators. This platform will serve as the foundation to develop a few emblematic applications. Several privacy-preserving applications can actually be targeted by a Personal Cloud platform, like: (i) smart disclosure applications allowing the individual to recover her personal data from external sources (e.g., bank, online shopping activity, insurance, etc.), integrate them and cross them to perform personal big data tasks (e.g., to improve her budget management) ; (ii) management of personal medical records for care coordination and well-being improvement; (iii) privacy-aware data management for the IoT (e.g., in sensors, quantified-self devices, smart meters); (iv) community-based sensing and community data sharing; (v) privacy-preserving studies (e.g., cohorts, public surveys, privacy-preserving data publishing). Such applications overlap with all the research axes described above but each of them also presents its own specificities. For instance, the smart disclosure applications will focus primarily on sharing models and enforcement, the IoT applications require to look with priority at the embedded data management and sustainability issues, while community-based sensing and privacy-preserving studies demand to study secure and efficient global query processing. Among these applications domains, one is already receiving a particular attention from our team. Indeed, we gained a strong expertise in the management and protection of healthcare data through our past DMSP (Dossier Medico-Social Partagé) experiment in the field. This expertise is being exploited to develop a dedicated healthcare and well-being personal cloud platform. We are currently deploying 10000 boxes equipped with PlugDB in the context of the DomYcile project. In this context, we are currently setting up an Inria Innovation Lab with the Hippocad company to industrialize this platform and deploy it at large scale (see Section the bilateral contract OwnCare II-Lab).

POTIOC Project-Team

4. Application Domains

4.1. Education, popularization of science, art, entertainment

Our project aims at providing rich interaction experiences between users and the digital world, in particular for non-expert users. The final goal is to stimulate understanding, learning, communication and creation. Our scope of applications encompasses

- education
- popularization of science
- art
- entertainment

See "Objective 3: Exploring new applications and usages" (3.4) for a detailed description.

<p style="text-align: center; color: red;">**RITS Project-Team**</p>

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. 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. RITS 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 RITS 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 to 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 EVA team for setting and tuning OLSR, a dynamic routing protocol for vehicles communications. 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.

<p align="center" style="color:red">SEMAGRAMME Project-Team</p>

4. Application Domains

4.1. Deep Semantic Analysis

Our applicative domains concern natural language processing applications that rely on a deep semantic analysis. For instance, one may cite the following ones:

- textual entailment and inference,
- dialogue systems,
- semantic-oriented query systems,
- content analysis of unstructured documents,
- text transformation and automatic summarization,
- (semi) automatic knowledge acquisition.

It seems clear, nowadays, that the need for semantics is ubiquitous. Nevertheless, according to the present state of the art, there are only a few applications for which a deep semantic analysis results in a real improvement over non semantic-based techniques. This is due to the fact that most current application chains are such that their weakest links are not located at the semantic level.

4.2. Text Transformation

Text transformation is an application domain featuring two important sub-fields of computational linguistics:

- parsing, from surface form to abstract representation,
- generation, from abstract representation to surface form.

Text simplification or automatic summarization belong to that domain.

We aim at using the framework of Abstract Categorial Grammars we develop to this end. It is indeed a reversible framework that allows both parsing and generation. Its underlying mathematical structure of λ-calculus makes it fit with our type-theoretic approach to discourse dynamics modeling.

SIROCCO Project-Team

4. Application Domains

4.1. Overview

The application domains addressed by the project are:

- Compression with advanced functionalities of various imaging modalities
- Networked multimedia applications taking into account needs in terms of user and network adaptation (e.g., interactive streaming, resilience to channel noise)
- Content editing, post-production, and computational photography.

4.2. Compression of emerging imaging modalities

Compression of visual content remains a widely-sought capability for a large number of applications. This is particularly true for mobile applications, as the need for wireless transmission capacity will significantly increase during the years to come. Hence, efficient compression tools are required to satisfy the trend towards mobile access to larger image resolutions and higher quality. A new impulse to research in video compression is also brought by the emergence of new formats beyond High Definition TV (HDTV) towards high dynamic range (higher bit depth, extended colorimetric space), or of formats for immersive displays allowing panoramic viewing, Free Viewpoint Video (FVV) and 3DTV.

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A multi-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FTV). In parallel, the advent of a variety of heterogeneous delivery infrastructures has given momentum to extensive work on optimizing the end-to-end delivery QoS (Quality of Service). This encompasses compression capability but also capability for adapting the compressed streams to varying network conditions. The scalability of the video content compressed representation and its robustness to transmission impairments are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

4.3. Networked visual applications

The emergence of multi-view auto-stereoscopic displays has spurred a recent interest for broadcast or Internet delivery of 3D video to the home. Multiview video, with the help of depth information on the scene, allows scene rendering on immersive stereo or auto-stereoscopic displays for 3DTV applications. This application sector suffers from an accommodation-vergence conflict which arises with conventional 3D displays (with or without glasses). Since each eye receives a single view, the eyes tend to focus on the display screen (accommodation), whereas the brain perceives the depth of 3D images due to the different views seen by each eye (vergence).

On the other hand, Free-viewpoint television (FTV) is a system for watching videos in which the user can choose its viewpoint freely and change it at anytime. To allow this navigation, many views are proposed and the user can navigate from one to the other. The goal of FTV is to propose an immersive sensation without the disadvantage of Three-dimensional television (3DTV). With FTV, a look-around effect is produced without any visual fatigue since the displayed images remain 2D. However, technical characteristics of FTV are large databases, huge numbers of users, and requests of subsets of the data, while the subset can be randomly chosen by the viewer. This requires the design of coding algorithms allowing such a random access to the pre-encoded and stored data which would preserve the compression performance of predictive coding. This research also finds applications in the context of Internet of Things in which the problem arises of optimally selecting both the number and the position of reference sensors and of compressing the captured data to be shared among a high number of users.

Broadband fixed (ADSL, ADSL2+) and mobile access networks with different radio access technologies (RAT) (e.g. 3G/4G, GERAN, UTRAN, DVB-H), have enabled not only IPTV and Internet TV but also the emergence of mobile TV and mobile devices with internet capability. A major challenge for next internet TV or internet video remains to be able to deliver the increasing variety of media (including more and more bandwidth demanding media) with a sufficient end-to-end QoS (Quality of Service) and QoE (Quality of Experience).

4.4. Editing, post-production and computational photography

Editing and post-production are critical aspects in the audio-visual production process. Increased ways of "consuming" visual content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. Content repurposing encompasses format conversion (retargeting), content summarization, and content editing. This processing requires powerful methods for extracting condensed video representations as well as powerful inpainting techniques. By providing advanced models, advanced video processing and image analysis tools, more visual effects, with more realism become possible. Our activies around light field imaging also find applications in computational photography which refers to the capability of creating photographic functionalities beyond what is possible with traditional cameras and processing tools.

STARS Project-Team

4. Application Domains

4.1. Introduction

While in our research the focus is to develop techniques, models and platforms that are generic and reusable, we also make effort in the development of real applications. The motivation is twofold. The first is to validate the new ideas and approaches we introduce. The second is to demonstrate how to build working systems for real applications of various domains based on the techniques and tools developed. Indeed, Stars focuses on two main domains: **video analytic** and **healthcare monitoring**.

4.2. Video Analytics

Stars has long-lasting experience in video analytics for security, an area of paramount importance due to the high need for public security, accompanied by the availability of a vast amount of cameras. Related applications range from public safety, smart retail, crowd management to maintenance of secure sites. One current challenge of video analytics is extracting the meaningful information from big video data in an efficient manner, with the goal of high-level semantic descriptions of the activities occurring in the area under surveillance.

Video analytics covers research directions including tracking, object detection, object classification, behavior recognition, pose estimation, and semantic scene modeling. Deep neural networks have brought to the fore great progress in the area of large-scale video analytics and specifically in object recognition and action recognition. However, many challenges remain open, mainly in the context of real-life surveillance scenarios and specifically concerning the robustness, adaptability, and scalability. Prominent open challenges comprise people detection and tracking in crowded environments, as well as detecting abnormal activities, which is often associated with processing of very large video data in the presence of noise and occlusions.

Stars has accumulated a well selected network of industrial partners ranging from end-users, integrators and software editors to provide data, objectives, evaluation and funding.

4.3. Healthcare Monitoring

Since 2011, we have initiated a strategic partnership (called CobTek) with Nice hospital [62], [112] (CHU Nice, Prof P. Robert) to start ambitious research activities dedicated to healthcare monitoring and to assistive technologies. These new studies address the analysis of more complex spatio-temporal activities (e.g. complex interactions, long term activities).

4.3.1. Research

To achieve this objective, several topics need to be tackled. These topics can be summarized within two points: finer activity description and longitudinal experimentation. Finer activity description is needed for instance, to discriminate the activities (e.g. sitting, walking, eating) of Alzheimer patients from the ones of healthy older people. It is essential to be able to pre-diagnose dementia and to provide a better and more specialized care. Longer analysis is required when people monitoring aims at measuring the evolution of patient behavioral disorders. Setting up such long experimentation with dementia people has never been tried before but is necessary to have real-world validation. This is one of the challenge of the European FP7 project Dem@Care where several patient homes should be monitored over several months.

For this domain, a goal for Stars is to allow people with dementia to continue living in a self-sufficient manner in their own homes or residential centers, away from a hospital, as well as to allow clinicians and caregivers remotely provide effective care and management. For all this to become possible, comprehensive monitoring of the daily life of the person with dementia is deemed necessary, since caregivers and clinicians will need a comprehensive view of the person's daily activities, behavioral patterns, lifestyle, as well as changes in them, indicating the progression of their condition.

4.3.2. Ethical and Acceptability Issues

The development and ultimate use of novel assistive technologies by a vulnerable user group such as individuals with dementia, and the assessment methodologies planned by Stars are not free of ethical, or even legal concerns, even if many studies have shown how these Information and Communication Technologies (ICT) can be useful and well accepted by older people with or without impairments. Thus one goal of Stars team is to design the right technologies that can provide the appropriate information to the medical carers while preserving people privacy. Moreover, Stars will pay particular attention to ethical, acceptability, legal and privacy concerns that may arise, addressing them in a professional way following the corresponding established EU and national laws and regulations, especially when outside France. Now, Stars can benefit from the support of the COERLE (Comité Opérationnel d'Evaluation des Risques Légaux et Ethiques) to help it to respect ethical policies in its applications.

As presented in 3.1 , Stars aims at designing cognitive vision systems with perceptual capabilities to monitor efficiently people activities. As a matter of fact, vision sensors can be seen as intrusive ones, even if no images are acquired or transmitted (only meta-data describing activities need to be collected). Therefore new communication paradigms and other sensors (e.g. accelerometers, RFID, and new sensors to come in the future) are also envisaged to provide the most appropriate services to the observed people, while preserving their privacy. To better understand ethical issues, Stars members are already involved in several ethical organizations. For instance, F. Brémond has been a member of the ODEGAM - "Commission Ethique et Droit" (a local association in Nice area for ethical issues related to older people) from 2010 to 2011 and a member of the French scientific council for the national seminar on "La maladie d'Alzheimer et les nouvelles technologies - Enjeux éthiques et questions de société" in 2011. This council has in particular proposed a chart and guidelines for conducting researches with dementia patients.

For addressing the acceptability issues, focus groups and HMI (Human Machine Interaction) experts, will be consulted on the most adequate range of mechanisms to interact and display information to older people.

THOTH Project-Team

4. Application Domains

4.1. Visual applications

Any solution to automatically understanding images and videos on a semantic level will have an immediate impact on a wide range of applications. For example:

- Semantic-level image and video access is highly relevant for visual search on the Web, in professional archives and personal collections.

- Visual data organization is applicable to organizing family photo and video albums as well as to large-scale information retrieval.

- Visual object recognition has potential applications ranging from surveillance, service robotics for assistance in day-to-day activities as well as the medical domain.

- Action recognition is highly relevant to visual surveillance, assisted driving and video access.

- Real-time scene understanding is relevant for human interaction through devices such as HoloLens, Oculus Rift.

4.2. Pluri-disciplinary research

Machine learning is intrinsically pluri-disciplinary. By developing large-scale machine learning models and algorithms for processing data, the Thoth team became naturally involved in pluri-disciplinary collaborations that go beyond visual modelling. In particular,

- extensions of unsupervised learning techniques originally developed for modelling the statistics of natural images have been deployed in neuro-imaging for fMRI data with the collaboration of the Parietal team from Inria.

- similarly, deep convolutional data representations, also originally developed for visual data, have been successfully extended to the processing of biological sequences, with collaborators from bio-informatics.

- Thoth also collaborates with experts in natural language and text processing, for applications where visual modalities need to be combined with text data.

TITANE Project-Team

4. Application Domains

4.1. Applications

In addition to tackling enduring scientific challenges, our research on geometric modeling and processing is motivated by applications to computational engineering, reverse engineering, digital mapping and urban planning. The main deliverable of our research will be algorithms with theoretical foundations. Ultimately we wish to contribute making geometry modeling and processing routine for practitioners who deal with real-world data. Our contributions may also be used as a sound basis for future software and technology developments.

Our first ambition for technology transfer is to consolidate the components of our research experiments in the form of new software components for the CGAL (Computational Geometry Algorithms Library) library. Consolidation being best achieved with the help of an engineer, we will search for additional funding. Through CGAL we wish to contribute to the "standard geometric toolbox", so as to provide a generic answer to application needs instead of fragmenting our contributions. We already cooperate with the Inria spin-off company Geometry Factory, which commercializes CGAL, maintains it and provide technical support.

Our second ambition is to increase the research momentum of companies through advising Cifre Ph.D. theses and postdoctoral fellows on topics that match our research program.

4. Application Domains

4.1. Web Programming Technologies

Despite the major social and economic impacts of the web revolution, current web programming methods and content representation are lagging behind and remain severely limited and in many respects archaic. Dangerously, designing web applications even becomes increasingly complex as it relies more and more on a jungle of programming languages, tools and data formats, each targeted toward a different application layer (presentation, application and storage). This often yields complex and opaque applications organized in silos, which are costly, inefficient, hard to maintain and evolve, and vulnerable to errors and security holes. In addition, the communication aspects are often handled independently via remote service invocations and represent another source of complexity and vulnerability. We believe that we reached a level where there is an urgent need and a growing demand for alternative programming frameworks that capture the essence of web applications: advanced content, data and communication. Therefore, successful candidate frameworks must capture rich document formats, data models and communication patterns. A crucial aspect is to offer correction guarantees and flexibility in the application architecture. For instance, applications need to be checked, optimized and managed as a whole while leveraging on the consistency of their individual components and data fragments. For all these reasons, we believe that a new generation of tools must be created and developed in order to overcome the aforementioned limitations of current web technologies.

4.2. Multimedia and Augmented Environments

The term Augmented Environments refers collectively to ubiquitous computing, context-aware computing, and intelligent environments. The goal of our research on these environments is to introduce personal Augmented Reality (AR) devices, taking advantage of their embedded sensors. We believe that personal AR devices such as mobile phones or tablets will play a central role in augmented environments. These environments offer the possibility of using ubiquitous computation, communication, and sensing to enable the presentation of context-sensitive information and services to the user. AR applications often rely on 3D content and employ specialized hardware and computer vision techniques for both tracking and scene reconstruction and exploration. Our approach tries to seek a balance between these traditional AR contexts and what has come to be known as mobile AR browsing. It first acknowledges that mobile augmented environment browsing does not require that 3D content be the primary means of authoring. It provides instead a method for HTML5 and audio content to be authored, positioned in the surrounding environments and manipulated as freely as in modern web browsers. The applications we develop to guide and validate our concepts are pedestrian navigation techniques and applications for cultural heritage visits. Features found in augmented environments are demanding for the other activities in the team. They require all kinds of multimedia information, that they have to combine. This information has to be processed efficiently and safely, often in real time, and it also, for a significant part, has to be created by human users.

4. Application Domains

4.1. Personal Information Management Systems

We recall that Valda's focus is on human-centric data, i.e., data produced by humans, explicitly or implicitly, or more generally containing information about humans. Quite naturally, we will use as a privileged application area to validate Valda's results that of personal information management systems (Pims for short) [1].

A Pims is a system that allows a user to integrate her own data, e.g., emails and other kinds of messages, calendar, contacts, web search, social network, travel information, work projects, etc. Such information is commonly spread across different services. The goal is to give back to a user the control on her information, allowing her to formulate queries such as "What kind of interaction did I have recently with Alice B.?", "Where were my last ten business trips, and who helped me plan them?". The system has to orchestrate queries to the various services (which means knowing the existence of these services, and how to interact with them), integrate information from them (which means having data models for this information and its representation in the services), e.g., align a GPS location of the user to a business address or place mentioned in an email, or an event in a calendar to some event in a Web search. This information must be accessed intensionally: for instance, costly information extraction tools should only be run on emails which seem relevant, perhaps identified by a less costly cursory analysis (this means, in turn, obtaining a cost model for access to the different services). Impacted people can be found by examining events in the user's calendar and determining who is likely to attend them, perhaps based on email exchanges or former events' participant lists. Of course, uncertainty has to be maintained along the entire process, and provenance information is needed to explain query results to the user (e.g., indicate which meetings and trips are relevant to each person of the output). Knowledge about services, their data models, their costs, need either to be provided by the system designer, or to be automatically learned from interaction with these services, as in [70].

One motivation for that choice is that Pims concentrate many of the problems we intend to investigate: heterogeneity (various sources, each with a different structure), massive distribution (information spread out over the Web, in numerous sources), rapid evolution (new data regularly added), intensionality (knowledge from Wikidata, OpenStreetMap...), confidentiality and security (mostly private data), and uncertainty (very variable quality). Though the data is distributed, its size is relatively modest; other applications may be considered for works focusing on processing data at large scale, which is a potential research direction within Valda, though not our main focus. Another strong motivation for the choice of Pims as application domain is the importance of this application from a societal viewpoint.

A Pims is essentially a system built on top of a user's *personal knowledge base*; such knowledge bases are reminiscent of those found in the Semantic Web, e.g., linked open data. Some issues, such as ontology alignment [73] exist in both scenarios. However, there are some fundamental differences in building personal knowledge bases vs collecting information from the Semantic Web: first, the scope is quite smaller, as one is only interested in knowledge related to a given individual; second, a small proportion of the data is already present in the form of semantic information, most needs to be extracted and annotated through appropriate wrappers and enrichers; third, though the linked open data is meant to be read-only, the only update possible to a user being adding new triples, a personal knowledge base is very much something that a user needs to be able to edit, and propagating updates from the knowledge base to original data sources is a challenge in itself.

4.2. Web Data

The choice of Pims is not exclusive. We intend to consider other application areas as well. In particular, we have worked in the past and have a strong expertise on Web data [3] in a broad sense: semi-structured, structured, or unstructured content extracted from Web databases [70]; knowledge bases from the Semantic Web

[73]; social networks [9]; Web archives and Web crawls [52]; Web applications and deep Web databases [45]; crowdsourcing platforms [40]. We intend to continue using Web data as a natural application domain for the research within Valda when relevant. For instance [44], deep Web databases are a natural application scenario for intensional data management issues: determining if a deep Web database contains some information requires optimizing the number of costly requests to that database.

A common aspect of both personal information and Web data is that their exploitation raises ethical considerations. Thus, a user needs to remain fully in control of the usage that is made of her personal information; a search engine or recommender system that ranks Web content for display to a specific user needs to do so in an unbiased, justifiable, manner. These ethical constraints sometimes forbid some technically solutions that may be technically useful, such as sharing a model learned from the personal data of a user to another user, or using blackboxes to rank query result. We fully intend to consider these ethical considerations within Valda. One of the main goals of a Pims is indeed to empower the user with a full control on the use of this data.

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. 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.

<p style="text-align:center; color:red; font-weight:bold;">WIMMICS Project-Team</p>

4. Application Domains

4.1. Social Semantic Web

A number of evolutions have changed the face of information systems in the past decade but the advent of the Web is unquestionably a major one and it is here to stay. From an initial wide-spread perception of a public documentary system, the Web as an object turned into a social virtual space and, as a technology, grew as an application design paradigm (services, data formats, query languages, scripting, interfaces, reasoning, etc.). The universal deployment and support of its standards led the Web to take over nearly all of our information systems. As the Web continues to evolve, our information systems are evolving with it.

Today in organizations, not only almost every internal information system is a Web application, but these applications also more and more often interact with external Web applications. The complexity and coupling of these Web-based information systems call for specification methods and engineering tools. From capturing the needs of users to deploying a usable solution, there are many steps involving computer science specialists and non-specialists.

We defend the idea of relying on Semantic Web formalisms to capture and reason on the models of these information systems supporting the design, evolution, interoperability and reuse of the models and their data as well as the workflows and the processing.

4.2. Linked Data on the Web and on Intranets

With billions of triples online (see Linked Open Data initiative), the Semantic Web is providing and linking open data at a growing pace and publishing and interlinking the semantics of their schemas. Information systems can now tap into and contribute to this Web of data, pulling and integrating data on demand. Many organisations also started to use this approach on their intranets leading to what is called linked enterprise data.

A first application domain for us is the publication and linking of data and their schemas through Web architectures. Our results provide software platforms to publish and query data and their schemas, to enrich these data in particular by reasoning on their schemas, to control their access and licenses, to assist the workflows that exploit them, to support the use of distributed datasets, to assist the browsing and visualization of data, etc.

Examples of collaboration and applied projects include: SMILK Joint Laboratory, Corese, DBpedia.fr.

4.3. Assisting Web-based Epistemic Communities

In parallel to linked open data on the Web, social Web applications also spread virally (e.g. Facebook growing toward 1.5 billion users) first giving the Web back its status of a social read-write media and then putting it back on track to its full potential of a virtual place where to act, react and interact. In addition, many organizations are now considering deploying social Web applications internally to foster community building, expert cartography, business intelligence, technological watch and knowledge sharing in general.

By reasoning on the Linked Data and the semantics of the schemas used to represent social structures and Web resources, we provide applications supporting communities of practice and interest and fostering their interactions in many different contexts (e-learning, business intelligence, technical watch, etc.).

We use typed graphs to capture and mix: social networks with the kinds of relationships and the descriptions of the persons; compositions of Web services with types of inputs and outputs; links between documents with their genre and topics; hierarchies of classes, thesauri, ontologies and folksonomies; recorded traces and suggested navigation courses; submitted queries and detected frequent patterns; timelines and workflows; etc.

Our results assist epistemic communities in their daily activities such as biologists exchanging results, business intelligence and technological watch networks informing companies, engineers interacting on a project, conference attendees, students following the same course, tourists visiting a region, mobile experts on the field, etc. Examples of collaboration and applied projects: EduMICS, OCKTOPUS, Vigiglobe, Educlever, Gayatech.

4.4. Linked Data for a Web of diversity

We intend to build on our results on explanations (provenance, traceability, justifications) and to continue our work on opinions and arguments mining toward the global analysis of controversies and online debates. One result would be to provide new search results encompassing the diversity of viewpoints and providing indicators supporting opinion and decision making and ultimately a Web of trust. Trust indicators may require collaborations with teams specialized in data certification, cryptography, signature, security services and protocols, etc. and this will raise the specific problem of interaction design for security and privacy. In addition, from the point of view of the content, this requires to foster the publication and coexistence of heterogeneous data with different points of views and conceptualizations of the world. We intend to pursue the extension of formalisms to allow different representations of the world to co-exist and be linked and we will pay special attention to the cultural domain and the digital humanities. Examples of collaboration and applied projects: Zoomathia, Seempad, SMILK,

4.5. Artificial Web intelligence

We intend to build on our experience in artificial intelligence (knowledge representation, reasoning) and distributed artificial intelligence (multi-agent systems - MAS) to enrich formalisms and propose alternative types of reasoning (graph-based operations, reasoning with uncertainty, inductive reasoning, non-monotonic, etc.) and alternative architectures for linked data with adequate changes and extensions required by the open nature of the Web. There is a clear renewed interest in AI for the Web in general and for Web intelligence in particular. Moreover distributed AI and MAS provide both new architectures and new simulation platforms for the Web. At the macro level, the evolution accelerated with HTML5 toward Web pages as full applications and direct Page2Page communication between browser clearly is a new area for MAS and P2P architectures. Interesting scenarios include the support to a strong decentralization of the Web and its resilience to degraded technical conditions (downscaling the Web), allowing pages to connect in a decentralized way, forming a neutral space, and possibly going offline and online again in erratic ways. At the micro level one can imagine the place RDF and SPARQL could take as data model and programming model in the virtual machines of these new Web pages and, of course, in the Web servers. RDF is also used to serialize and encapsulate other languages and becomes a pivot language in linking very different applications and aspects of applications. Example of collaboration and applied projects: MoreWAIS, Corese, Vigiglobe collaboration.

4.6. Human-Data Interaction (HDI) on the Web

We need more interaction design tools and methods for linked data access and contribution. We intend to extend our work on exploratory search coupling it with visual analytics to assist sense making. It could be a continuation of the Gephi extension we built targeting more support for non expert to access and analyze data on a topic or issue of their choice. More generally speaking SPARQL is inappropriate for common users and we need to support a larger variety of interaction means with linked data. We also believe linked data and natural language processing (NLP) have to be strongly integrated to support natural language based interactions. Linked Open Data (LOD) for NLP, NLP for LOD and Natural Dialog Processing for querying, extracting and asserting data on the Web is a priority to democratize its use. Micro accesses and micro contributions are important to ensure public participation and also call for customized interfaces and thus for methods and tools to generate these interfaces. In addition, the user profiles are being enriched now with new data about the user such as his current mental and physical state, the emotion he just expressed or his cognitive performances. Taking into account this information to improve the interactions, change the behavior of the system and adapt the interface is a promising direction. And these human-data interaction means should

also be available for "small data", helping the user to manage her personal information and to link it to public one or collective one maintaining her personal and private perspective as a personal Web of data. Finally, the continuous knowledge extractions, updates and flows add the additional problem of representing, storing, querying and interacting with dynamic data. Examples of collaboration and applied projects: QAKIS, Sychonext collaboration, ALOOF, DiscoveryHub, Wasabi, MoreWAIS.

Web-augmented interactions with the world: The Web continues to augment our perception and interaction with reality. In particular, Linked Open Data enable new augmented reality applications by providing data sources on almost any topic. The current enthusiasm for the Web of Things, where every object has a corresponding Web resource, requires evolutions of our vision and use of the Web architecture. This vision requires new techniques as the ones mentioned above to support local search and contextual access to local resources but also new methods and tools to design Web-based human devices interactions. accessibility, etc.) These new usages are placing new requirements on the Web Architecture in general and on the semantic Web models and algorithms in particular to handle new types of linked data. They should support implicit requests considering the user context as a permanent query. They should also simplify our interactions with devices around us jointly using our personal preferences and public common knowledge to focus the interaction on the vital minimum that cannot be derived in another way. For instance the access to the Web of data for a robot can completely change the quality of the interactions it can offer. Again these interactions and the data they require raise problems of security and privacy. Examples of collaboration and applied projects: ALOOF, AZKAR, MoreWAIS.

ZENITH Project-Team

4. Application Domains

4.1. Data-intensive Scientific Applications

The application domains covered by Zenith are very wide and diverse, as they concern data-intensive scientific applications, i.e., most scientific applications. Since the interaction with scientists is crucial to identify and tackle data management problems, we are dealing primarily with application domains for which Montpellier has an excellent track record, i.e., agronomy, environmental science, life science, with scientific partners like INRA, IRD and CIRAD. However, we are also addressing other scientific domains (e.g. astronomy, oil extraction) through our international collaborations (e.g. in Brazil).

Let us briefly illustrate some representative examples of scientific applications on which we have been working on.

- **Management of astronomical catalogs**. An example of data-intensive scientific applications is the management of astronomical catalogs generated by the Dark Energy Survey (DES) project on which we are collaborating with researchers from Brazil. In this project, huge tables with billions of tuples and hundreds of attributes (corresponding to dimensions, mainly double precision real numbers) store the collected sky data. Data are appended to the catalog database as new observations are performed and the resulting database size is estimated to reach 100TB very soon. Scientists around the globe can query the database with queries that may contain a considerable number of attributes. The volume of data that this application holds poses important challenges for data management. In particular, efficient solutions are needed to partition and distribute the data in several servers. An efficient partitioning scheme should try to minimize the number of fragments accessed in the execution of a query, thus reducing the overhead associated to handle the distributed execution.

- **Personal health data analysis and privacy** Today, it is possible to acquire data on many domains related to personal data. For instance, one can collect data on her daily activities, habits or health. It is also possible to measure performance in sports. This can be done thanks to sensors, communicating devices or even connected glasses. Such data, once acquired, can lead to valuable knowledge for these domains. For people having a specific disease, it might be important to know if they belong to a specific category that needs particular care. For an individual, it can be interesting to find a category that corresponds to her performances in a specific sport and then adapt her training with an adequate program. Meanwhile, for privacy reasons, people will be reluctant to share their personal data and make them public. Therefore, it is important to provide them solutions that can extract such knowledge from everybody's data, while guaranteeing that their private data won't be disclosed to anyone.

- **Botanical data sharing**. Botanical data is highly decentralized and heterogeneous. Each actor has its own expertise domain, hosts its own data, and describes them in a specific format. Furthermore, botanical data is complex. A single plant's observation might include many structured and unstructured tags, several images of different organs, some empirical measurements and a few other contextual data (time, location, author, etc.). A noticeable consequence is that simply identifying plant species is often a very difficult task; even for the botanists themselves (the so-called taxonomic gap). Botanical data sharing should thus speed up the integration of raw observation data, while providing users an easy and efficient access to integrated data. This requires to deal with social-based data integration and sharing, massive data analysis and scalable content-based information retrieval. We address this application in the context of the French initiative Pl@ntNet, with CIRAD and IRD.

- **Biology data integration and analysis**.

Biology and its applications, from medicine to agronomy and ecology, are now producing massive data, which is revolutionizing the way life scientists work. For instance, using plant phenotyping platforms such as PhenoDyn at INRA Montpellier, quantitative genetic methods allow to identify genes involved in phenotypic variation in response to environmental conditions. These methods produce large amounts of data at different time intervals (minutes to days), at different sites and at different scales ranging from small tissue samples until the entire plant. Analyzing such big data creates new challenges for data management and data integration.

These application examples illustrate the diversity of requirements and issues which we are addressing with our scientific application partners. To further validate our solutions and extend the scope of our results, we also want to foster industrial collaborations, even in non scientific applications, provided that they exhibit similar challenges.