Activity Report 2014

Section Application Domains

PERCEPTION, COGNITION AND INTERACTION

ANTIQUE 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, ASTRÉEsuccessfully 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.

AOSTE Project-Team

4. Application Domains

4.1. Multicore System-on-Chip design

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

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

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

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

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

4.2. Automotive and avionic embedded systems

Model-Driven Engineering is in general well accepted in the transportation domains, where design of digital software and electronic parts in usually tighly coupled with larger aspects of system design, where models from physics are being used already. The formalisms AADL (for avionics) and AutoSar [55] (for automotive) are providing support for this, unfortunately not always with a clean and formal semantics. Thus there is a strong need here for approaches that bring closer together formal methods and tools on the one hand, engineering best practices on the other hand.

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

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

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

<div align="center">

CASCADE Project-Team

</div>

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.

CRYPT Team (section vide)

<div align="center">**DEDUCTEAM Exploratory Action**</div>

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. B-set theory

Set theory appears to be an appropriate theory for automated theorem provers based on Deduction modulo, in particular the several extensions of *Zenon* (*Super Zenon* and *Zenon Modulo*). Modeling techniques using set theory are therefore good candidates to assess these tools. This is what we have done with the *B* method whose formalism relies on set theory. A collaboration with *Siemens* has been developed to automatically verify the *B* proof rules of *Atelier B* [42]. From this work presented in the Doctoral dissertation of Mélanie Jacquel, the *Super Zenon* tool [13] [35] has been designed in order to be able to reason modulo the *B* set theory. As a sequel of this work, we contribute to the *BWare* project whose aim is to provide a mechanized framework to support the automated verification of *B* proof obligations coming from the development of industrial applications. In this context, we have recently designed *Zenon Modulo* [33], [34] (Pierre Halmagrand's PhD thesis, which has started on October 2013) to deal with the *B* set theory. In this work, the idea is to manually transform the *B* set theory into a theory modulo and provide it to *Zenon Modulo* in order to verify the proof obligations of the *BWare* project.

4.3. 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 since several years now. Deducteam develops a formally certified tool, Rainbow, based on the Coq library CoLoR, that is able to automatically verify the correctness of such termination certificates.

GALLIUM Project-Team

4. Application Domains

4.1. High-assurance software

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

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

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

4.2. Software security

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

The methods used in designing type systems and establishing their soundness can also deliver static analyses that automatically verify some security policies. Two examples from our past work include Java bytecode verification [62] and enforcement of data confidentiality through type-based inference of information flows and noninterference properties [67].

4.3. Processing of complex structured data

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

4.4. Rapid development

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

4.5. Teaching programming

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

MUTANT Project-Team

4. Application Domains

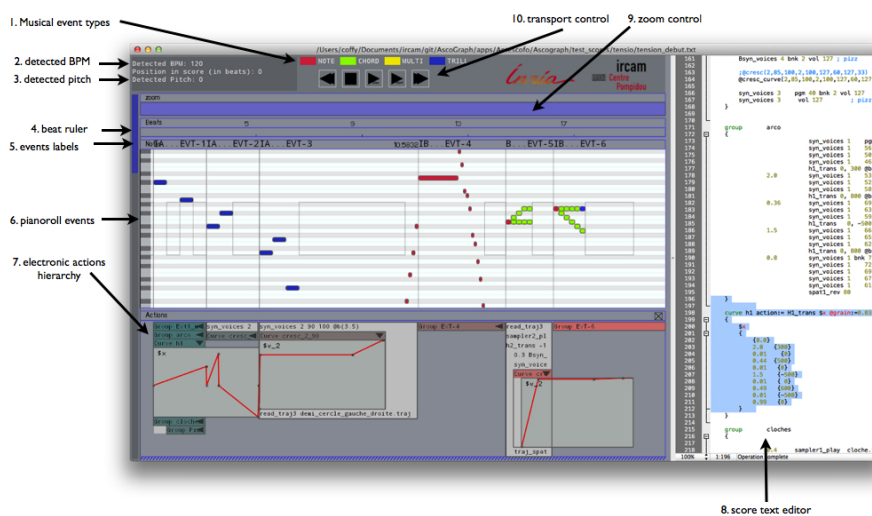## 4.1. Authoring and Performing Interactive Music

Figure 2. Screenshot of Ascograph, the Antescofo graphical score editor

The combination of both realtime machine listening systems and reactive programming paradigms has enabled the *authoring* of interactive music systems as well as their realtime performance within a coherent synchronous framework called Antescofo. The module, developed since 2008 by the team members, has gained increasing attention within the user community worldwide with more than 40 prestigious public performances yearly. The outcomes of the teams's research will enhance the interactive and reactive aspects of this emerging paradigm as well as creating novel authoring tool for such purposes.

The *AscoGraph* authoring environment, started in 2013 and shown in Figure 2 , is the first step towards such authoring environments. The outcome of the ANR Project INEDIT (with LABRI and GRAME and coordinated by team leader), will further extend the use-cases of *Antescofo* for interactive multimedia pieces with more complex temporal structures and computational paradigms.

4.2. Realtime Music Information Retrieval

Realtime Music Information Retrieval is used as front-end for various applications requiring sonic interaction between software/hardware and the physical worlds. MuTant has focused on realtime machine listening since its inception and holds state-of-the-art algorithms for realtime alignment of audio to symbolic score, realtime tempo detection, realtime multiple-pitch extraction. Recent results have pushed our application to more generalised listening schemes beyond music signals as reported in [26].

4.3. Automatic Accompaniment/Creative Tools for Entertainment Industry
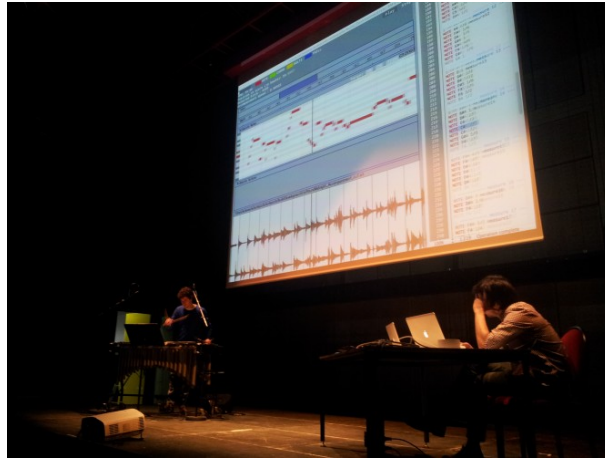
Figure 3. Automatic Accompaniment Session with Antescofo during ACM CHI 2013 Conference

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

Highlights in 2014 include collaborations with Orchestre de Paris Archives that resulted in prototype demonstrated to public in June 2014.

We will pursue this by licensing our technologies to third-party companies.

4. Application Domains

4.1. Provably safe and efficient computing systems

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

PI.R2 Project-Team (section vide)

4. Application Domains

4.1. Cryptology

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

4.2. Engineering sciences

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

<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.

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

4. Application Domains

4.1. Domain

Our main application domains are:

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

TEMPO Team

4. Application Domains

4.1. Cyber Physical Systems

The overall project is geared towards the development of complex heterogeneous cyber physical systems that require high reliability such as nuclear power, energy distribution, industry automation and transportation, where formal verification methods are necessary.

4.2. Simulation

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

4.3. Automated Test Generation

Manual testing is tedious. Automated testing makes it possible to increase test coverage while also minimizing the amount of redundancy created by manual testing.

CLASSIC Project-Team (section vide)

GAMMA3 Project-Team (section vide)

MATHERIALS Team

4. Application Domains

4.1. Homogenization and related problems

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

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

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

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

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

4.2. Electronic structure of large systems

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

- how can one improve the nonlinear iterations that are the basis of any *ab initio* models for computational chemistry?

- how can one more efficiently solve the inner loop which most often consists in the solution procedure for the linear problem (with frozen nonlinearity)?

- how can one design a sufficiently small variational space, whose dimension is kept limited while the size of the system increases?

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

- in the former, the results of the model at the lower scale are simply used to evaluate some parameters that are inserted in the model for the larger scale: one example is the parameterized classical molecular dynamics, which makes use of force fields that are fitted to calculations at the quantum level;

- while in the latter, the model at the lower scale is concurrently coupled to the model at the larger scale: an instance of such a strategy is the so called QM/MM coupling (standing for Quantum Mechanics/Molecular Mechanics coupling) where some part of the system (typically the reactive site of a protein) is modeled with quantum models, that therefore accounts for the change in the electronic structure and for the modification of chemical bonds, while the rest of the system (typically the inert part of a protein) is coarse grained and more crudely modeled by classical mechanics.

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

4.3. Computational Statistical Mechanics

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

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

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

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

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

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

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

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

4. Application Domains

4.1. Application Domains

Risk management, Quantitative finance, Computational Finance, Market Microstructure analysis, Systemic risk, Portfolio optimization, Risk modeling, Option pricing and hedging in incomplete markets, insurance.

MOKAPLAN Team

4. Application Domains

4.1. Continuous models in economics

- As already mentioned the CFD formulation is a limit case of simple variational Mean-Field Games (MFG) [65]. MFG is a new branch of game theory recently developed by J-M. Lasry and P-L. Lions. MFG models aim at describing the limiting behavior of stochastic differential games when the number of players tends to infinity. They are specifically designed to model economic problems where a large number of similar interacting agents try to maximize/minimize a utility/cost function which takes into account global but partial information on the game. The players in these models are individually insignificant but they collectively have a significant impact on the cost of the other players. Dynamic MFG models often lead to a system of PDEs which consists of a backward Hamilton-Jacobi Bellman equation for a value function coupled with a forward Fokker-Planck equation describing the space-time evolution of the density of agents.

- In microeconomics, the *principal-agent problem* [83] with adverse selection plays a distinguished role in the literature on asymmetric information and contract theory (with important contributions from several Nobel prizes such as Mirrlees, Myerson, Spence or Tirole) and it has many important applications in optimal taxation, insurance, nonlinear pricing. The problem can be reduced to the maximization of an integral functional subject to a convexity constraint This is an unusual calculus of variations problem and the optimal price can only be computed numerically. Recently, following a reformulation of Carlier [12], convexity/well-posedness results of McCann, Figalli and Kim [52], connected to optimal transport theory, showed that there is some hope to numerically solve the problem for general utility functions.

- In [9] a class of games are considered with a continuum of players for which Cournot-Nash equilibria can be obtained by the minimisation of some cost, related to optimal transport. This cost is not convex in the usual sense in general but it turns out to have hidden strict convexity properties in many relevant cases. This enables us to obtain new uniqueness results and a characterisation of equilibria in terms of some partial differential equations, a simple numerical scheme in dimension one as well as an analysis of the inefficiency of equilibria. The mathematical problem has the structure of one step of the JKO gradient flow method.

- Many relevant markets are markets of indivisible goods characterized by a certain quality: houses, jobs, marriages... On the theoretical side, recent papers by Ekeland, McCann, Chiappori [45] showed that finding equilibria in such markets is equivalent to solving a certain optimal transport problem (where the cost function depends on the sellers and buyers preferences). On the empirical side, this allows for trying to recover information on the preferences from observed matching; this is an inverse problem as in a recent work of Galichon and Salanié [57] [58] Interestingly, these problems naturally lead to numerically challenging variants of the Monge-Kantorovich problem: the multi-marginal OT problem and the entropic approximation of the Monge-Kantorovich problem (which is actually due to Schrödinger in the early 30's).

4.2. Finance

The Skorohod embedding problem (SEP) consists in finding a martingale interpolation between two probability measures. When a particular stochastic ordering between the two measures is given, Galichon et al [56] have shown that a very natural variational formulation could be given to a class of problems that includes the SEP. This formulation is related to the CFD formulation of the OT problem [2] and has applications to *model-free bounds of derivative prices in Finance*. It can also be interpreted as a a multi marginal Optimal Mass Transportation with infinitely many marginals [78].

4.3. Congested Crowd motion

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

- Optimal Mass Transportation and MFG theories can be an extremely powerful tool to attack some of these problems arising from spatial economics or to design new ones. For instance, various urban/traffic planning models have been proposed by Buttazzo, Santambrogio, Carlier ([10]   [40] [32]) in recent years.

- Many models from PDEs and fluid mechanics have been used to give a description of *people or vehicles moving in a congested environment*. These models have to be classified according to the dimension (1D model are mostly used for cars on traffic networks, while 2D models are most suitable for pedestrians), to the congestion effects ("soft" congestion standing for the phenomenon where high densities slow down the movement, "hard" congestion for the sudden effects when contacts occur, or a certain threshold is attained), and to the possible rationality of the agents Maury et al [69] recently developed a theory for 2D hard congestion models without rationality, first in a discrete and then in a continuous framework. This model produces a PDE that is difficult to attack with usual PDE methods, but has been successfully studied via Optimal Mass Transportation techniques again related to the JKO gradient flow paradigm.

4.4. Astrophysics

In [54] and [37], the authors show that the deterministic past history of the Universe can be uniquely reconstructed from the knowledge of the present mass density field, the latter being inferred from the 3D distribution of luminous matter, assumed to be tracing the distribution of dark matter up to a known bias. Reconstruction ceases to be unique below those scales – a few Mpc – where multi-streaming becomes significant. Above 6 Mpc/h we propose and implement an effective Monge-Ampere-Kantorovich method of unique reconstruction. At such scales the Zel'dovich approximation is well satisfied and reconstruction becomes an instance of optimal mass transportation. After discretization into N point masses one obtains an assignment problem that can be handled by effective algorithms with not more than cubic time complexity in N and reasonable CPU time requirements. Testing against N-body cosmological simulations gives over 60% of exactly reconstructed points.

4.5. Image Processing and inverse problems

The Wasserstein distance between densities is the value function of the Optimal Mass Transportation problem. This distance may be considered to have "orthogonal" properties to the widely used least square distance. It is for instance quadratic with respect to dilations and translation. On the other hand it is not very sensitive to rigid transformations, [75] is an attempt at generalizing the CFD formulation in this context. The Wasserstein distance is an interesting tool for applications where distances between signals and in particular oscillatory signals need to to computed, this is assuming one understands how to transform the information into positive densities.

- Tannenbaum and co-authors have designed several variants of the CFD numerical method and applied it to warping, morphing and registration (using the Optimal Mass Transportation map) problems in medical imaging. [86]   [30]

- Gabriel Peyre and co-authors [82] have proposed an easier to compute relaxation of the Wasserstein distance (the sliced Wasserstein distance) and applied it to two image processing problems: color transfer and texture mixing.

- Froese Engquist [51] use a Monge-Ampère Solver to compute the Wasserstein distance between synthetic 2D Seismic signals (After some transformations). Applications to waveform inversion and registration are discussed and simple numerical examples are presented.

4.6. Meteorology and Fluid models

In, [34] Brenier reviews in a unified framework the connection between optimal transport theory and classical convection theory for geophysical flows. Inspired by the numerical model proposed in [30], the starting point is a generalization of the Darcy-Boussinesq equations, which is a degenerate version of the Navier-Stokes-Boussinesq (NSB) equations. In a unified framework, he relates different variants of the NSB equations (in particular what he calls the generalized hydrostatic-Boussinesq equations) to various models involving optimal transport and the related Monge-Ampère equation. This includes the 2D semi-geostrophic equations [61] [49] [48] [4] [67] and some fully nonlinear versions of the so-called high-field limit of the Vlasov-Poisson system [73] and of the Keller-Segel system for chemotaxis [63] [44] .

4.7. Mesh motion/Lagragian methods

The necessity to preserve areas/volumes is a intrinsic feature of mesh deformations more generally Lagrangian numerical methods. Numerical method of Optimal Mass Transportation which preserve some notions of convexity and as a consequence the monotonicity of the computed transport maps can play a role in this context, see for instance [43] [46] [66].

4.8. Density Functionnal Theory (DFT)

The precise modeling of electron correlations continues to constitute the major obstacle in developing high-accuracy, low-cost methods for electronic structure computations in molecules and solids. The article [47] sheds a new light on the longstanding problem of how to accurately incorporate electron correlation into DFT, by deriving and analyzing the semiclassical limit of the exact Hohenberg-Kohn functional with the single-particle density ρ held fixed. In this limit, in the case of two electrons, the exact functional reduces to a very interesting functional that depends on an optimal transport map M associated with a given density ρ. The limit problem is known in the DFT literature with the optimal transport map being called a correlation function or a co-motion function , but it has not been rigorously derived, and it appears that it has not previously been interpreted as an optimal transport problem. The article [47] thereby links for the first time DFT, which is a large and very active research area in physics and chemistry, to optimal transportation theory with a Coulombian repulsive cost. Numerics are still widely open [38].

QUANTIC Team

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.

SIERRA Project-Team (section vide)

ANGE Project-Team

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, biogeochemistry of oceans, resilience of society w.r.t. hazardous flows,...

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 quasi-steady state, which is the balance between the gravity and the Coriolis forces, can significantly improve the forecasts. In addition, the 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 costly and generally not efficient in long term. The objective is to determine a configuration almost stable with the facilities. In addition, it is also important to determine the impact of major events like emptying dam. Iy 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 the damages by means of suitable amenities. In France, floods are the most recurring natural disasters and produce the worse damages. In addition, it can be a cause or a consequence of a dam break. The large surface they cover and the long period they can last requires 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 required fluid/structure interactions to be well understood. Moreover, underground flows, in particular in sewer, 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 not the case close to the epicenter and in the coastal zone where the bathymetry leads a vertical accretion and produce substantial dispersive effects. The non-hydrostatic terms have to be considered and an efficient numerical resolution should be induced.

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

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 CO_2 fixation is the subject of intense academic and industrial research. Large-scale production of micro-algae has potential for biofuel applications owing to the high productivity that can be attained in high-rate raceway ponds. One of the key challenges in the production of micro-algae is to maximize algae growth with respect to the exogenous energy that must be used (paddlewheel, pumps,...). There is a large number of parameters that need to be optimized (characteristics of the biological species, raceway shape, stirring provided by the paddlewheel). Consequently our strategy is to develop efficient models and numerical tools to reproduce the flow induced by the paddlewheel and the evolution of the biological species within this flow. Here, mathematical models can greatly help us reduce experimental costs. Owing to the high heterogeneity of raceways due to gradients of temperature, light intensity and nutrient availability through water height, we cannot use depth-averaged models. We adopt instead more accurate multilayer models that have recently been proposed. 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 a kinetic and potential part. The buoys use the potential energy whereas the turbines (hydrolian) 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.

ARAMIS Project-Team

4. Application Domains

4.1. Introduction

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

- better understanding the pathophysiology of brain disorders;
- designing biomarkers of pathologies for diagnosis, prognosis and assessment of drug efficacy;
- developping brain computer interfaces for clinical applications;
- improving the localisation of stimulation targets in Deep Brain Stimulation protocol.

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

4.2. Understanding brain disorders

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

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

4.3. Biomarkers for diagnosis, prognosis and clinical trials

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

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

4.4. Brain computer interfaces for clinical applications

A brain computer interface (BCI) is a device aiming to decode brain activity, thus creating an alternate communication channel between a person and the external environment. BCI systems can be categorized on the base of the classification of an induced or evoked brain activity. The central tenet of a BCI is the capability to distinguish different patterns of brain activity, each being associated to a particular intention or mental task. Hence adaptation, as well as learning, is a key component of a BCI because users must learn to modulate their brainwaves to generate distinct brain patterns. Usually, a BCI is considered a technology for people to substitute some lost functions. However, a BCI could also help in clinical rehabilitation to recover motor functions. Indeed, in current neuroscience-based rehabilitation it is recognized that protocols based on mental rehearsal of movements (like motor imagery practicing) are a way to access the motor system because they can induce an activation of sensorimotor networks that were affected by lesions. Hence, a BCI based on movement imagery can objectively monitor 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. The BCI approaches that we develop are based on the characterization of the information contained in the functional connectivity patterns. We expect to significantly increase the performance of the BCI system with respect to the sole use of standard power spectra of the activity generated by single local brain areas. Such an improvement will concretely provide the user with a more precise control of the external environment in open-loop BCI tasks and a more coherent feedback in the closed-loop BCI schemes.

4.5. Deep Brain Stimulation

Deep Brain Stimulation (DBS) is a surgical technique, which consists in sending electrical impulses, through implanted electrodes, to specific parts of the brain for the treatment of movement and affective disorders. The technique has been initially developped for otherwise-treatment-resistant patients with essential tremors or Parkinson's disease. Its benefit in other affections, such as dystonia, obsessive-compulsive disorders, Tourette syndrome is currently investigated. The localisation of the stimulation target in specific nucleus in deep brain regions is key to the success of the surgery. This task is difficult since the target nucleus, or the precise sub-territory of a given nucleus is rarely visible in the Magnetic Resonance Image (MRI) of the patients. To address this issue, a possible technique is to personalize a high-resolution histological atlas of the brain to each patient. This personalization is achieved by registering the histological atlas, which consists of an image and meshes of deep brain structures, to the pre-operative MRI of each patient. The registration is currently done by optimally aligning image intensities in the atlas and patient's MRI using a block-matching algorithm. The linear nature of the transform makes the technique robust at the cost of a lack of precision, especially for elderly patients with expanded ventricles. We investigate the use of non-linear registration techniques to optimally align both image intensities and contours of visible structures surrounding the target. We expect to improve the localisation of the target for patients with large ventricles while keeping the method robust in all cases.

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

4. Application Domains

4.1. Introduction

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

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

4.2. Air quality

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

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

- The development of ensemble forecast methods for estimating the quality of the prediction, in relation with the quality of the model and the observations. The ensemble methods allow sensitivity analysis with respect to the model's parameters so as to identify physical and chemical processes, whose modeling must be improved.

- The development of methodologies for sequential aggregation of ensemble simulations. What ensembles should be generated for that purpose, how spatialized forecasts can be generated with aggregation, how can the different approaches be coupled with data assimilation?

- The definition of second-order data assimilation methods for the design of optimal observation networks. The two main objectives are: management of combinations of sensor types and deployment modes and dynamic management of mobile sensors' trajectories.

- How to estimate the emission rate of an accidental release of a pollutant, using observations and a dispersion model (from the near-field to the continental scale)? How to optimally predict the evolution of a plume? Hence, how to help people in charge of risk evaluation for the population?

- The definition of non-Gaussian approaches for data assimilation.

- The assimilation of satellite measurements of troposphere chemistry.

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

4.3. Oceanography

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

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

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

4.4. Meteorology

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

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

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

LIFEWARE Team

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. We work mainly on eukaryotic cells. Our collaborations with biologists are focused on **concrete biological questions**, and on the building of predictive models of biological systems to answer them. Moreover, one important application of our research is the development of a **modeling platform** for systems biology.

4.2. Modeling platform 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 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 vesicles

In collaboration with Franck Molina (CNRS, Sysdiag, 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 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 Sysdiag, 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 is also particularly appealing for security considerations in synthetic biology 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.

4. Application Domains

4.1. Cancer modelling

Evolution of healthy or cancer cell populations under environmental pressure; drug resistance. Considering cancer as an *evolutionary disease* – evolution meaning here Darwinian evolution of populations structured according to relevant phenotypes – in collaboration with our biologist partners within the Institut Universitaire de Cancérologie (IUC) of UPMC, we tackle the problem of understanding and limiting a) evolution from pre-malignancy to malignancy in cell populations, and b) in established cancer cell populations, evolution towards (drug-induced) drug resistance. Environmental pressure guiding evolution is of various natures, including signalling molecules induced by the peritumoral stroma (e.g., between a breast tumour and its adipocytic stroma), and anticancer drugs and their effects on both the tumour and its stromal environment. The models we use [59], [58] [40] are akin to models used in ecology for adaptive dynamics.

Multi-scale modelling of EMT. The major step from a benign tumour that can be eradicated by surgery and an invasive cancer is the development step at which cells detach from the tumour mass and invade individually the surrounding tissue [0]. The invasion is preceded by a transition (called EMT - epithelial mesenchymal transition) of the cancer phenotype from an epithelial type to a mesenchymal type cell. We so far worked on multi-scale modelling of EMT [0], and the step by which invading cancer cells enter blood vessels, called intravasation [0]. We now perform in-vitro simulations of cancer cell invasion for Non Small Cell Lung Cancer (NSCLC) having a 5-year survival fraction of about 20%, and for breast cancer. Under development (in collaboration with our biologist partners within the IUC for the experimental part) is also a phenotype-structured PDE model of the interactions between colonies of MCF7 breast cancer and adipocyte stromal support populations.

4.2. Cancer therapies and their optimisation

Drugs. We focus on multi-drug multi-targeted anticancer therapies aiming at finding combinations of drugs that theoretically minimise cancer cell population growth with the constraint of limiting unwanted toxic side effects under an absolute threshold (i.e., this is not L^2 nor L^1, but L^∞ optimisation) in healthy cell populations and avoiding the emergence of resistant cell clones in cancer cell populations [49], [58] [4], [40]. Prior to using optimisation methods, we design models of the targeted cell populations (healthy and tumour, including molecular or functional drug targets [48]) by PDEs or agent-based models [3], and molecular pharmacological (pharmacokinetic-pharmacodynamic, PK-PD) models of the fate and effects of the drugs used, usually by ODE models. A particular aspect of such modelling is the representation of multi-cellular spatio-temporal patterns emerging from therapies.

Radiotherapy. Radiation is still a major treatment in cancer. We have recently published new results on this topic, pleae refer to the *New Results* section.

---

[0]Weinberg, The biology of cancer, Garland, 2007
[0]Ramis-Conde, Drasdo, Anderson, Chaplain, Biophys. J., 2008
[0]Ramis-Conde, Chaplain, Anderson, Drasdo, Phys. Biol. 2009

4.3. Cell motion

Several processes are employed by cells to communicate, regulate and control their movements and generate collective motion. Among them, chemotaxis is the phenomenon by which cells direct their active motion in response to an external chemical (or physical) agent. In chemotaxis, cells not only respond but can also produce the chemical agent, leading to a feedback loop. Understanding this phenomenon is a major challenge for describing the collective behaviour of cells. Many mathematical models have been proposed at different scales, yielding a good description of cell aggregation. In collaboration with biophysicists at Institut Curie in Paris, we develop and study [0] mathematical models based on kinetic equations for bacterial travelling waves in a microchannel. These models have shown a remarkable quantitative agreement with experimental observations.

Cell motion arises also in the growth of solid tumours, which can be described through cell population models or multiphase flows [0]. This is a very active subject because several bio-chemico-physical mechanisms are at work; for instance motion can arise from pressure forces resulting from cell divisions and from active cell motility. At the smaller scale stochastic agent-based models of tumour cells invading the tumour environment or blood vessels are considered [0], and allow to include detailed behaviours and interactions. At a larger scale, free boundary problems are widely used, e.g. for image-based prediction because of the reduced number of parameters [0]. Asymptotic analysis makes a link between these different mechanistic models [63].

One other setting where we will study cell motion is epithelial gap closure, a form of collective cell migration that is a very widespread phenomenon both during development and adult life - it is essential for both the formation and for the maintenance of epithelial layers. Due to their importance, wound healing in vivo and morphogenetic movements involving closure of holes in epithelia have been the object of many studies (including some involving members of this project like [47]). Several theoretical models have also been proposed recently for the advancement of tissue covering unoccupied areas (see, for instance, [46]). It is particularly interesting to study epithelial gap closure in vivo. However, the complexity of the process and the difficulty to measure relevant quantities directly and to control the parameters in vivo, lead people to seek alternative systems where epithelial gap closure can be studied under better-defined and better-controlled conditions.

4.4. Contraction of acto-myosin structures in morphogenesis and tissue repair

In 2014, L. Almeida, I. Cheddadi, C. Emako-Kazianou, P. Bagnerini [0], A. Jacinto [0], P. Patricio [0], B. Ladoux [0] and N. Gov [0] have continued to investigate the dependence of physical and biological mechanisms of actomyosin cable formation and wound closure depending on the geometry of the wound and adhesion to the substrate, with particular emphasis on the effect of the wound edge curvature. We extended our work from in vivo studies to in vitro situations taking advantage of a collaboration with the group of Benoît Ladoux who did experiments on cell monolayers of human keratinocytes and of MDCK cells. We could single out some similar geometry dependence of the wound closure strategies between these two settings indicating the existence of conserved mechanisms that should be very general across living beings.

In our model under development, we consider viscous behaviour and friction in the tissue plus boundary terms associated to cable and lamellipodial forces. The numerical simulations obtained using this model are in good agreement with the experimental results. This work is attracting considerable attention from the community.

---

[0]N. Bournaveas, V. Calvez, S. Gutiérrez and B. Perthame, Global existence for a kinetic model of chemotaxis via dispersion and Strichartz estimates, *Comm. PDE*, 2008

[0]J. Ranft et al, Fluidization of tissues by cell division and apoptosis, *PNAS*, 2010 and L. Preziosi and A. Tosin, Multiphase modelling of tumour growth and extracellular matrix interaction: mathematical tools and applications, *J. Math. Biol.*, 2009.

[0]I. Ramis-Conde et al., J. Phys. Biol., 2009

[0]Works by O. Saut, T. Colin, A. Iollo, N. Ayache, J. Lowengrub

[0]Univ. Genova

[0]CEDOC, Lisbon

[0]CEDOC, Lisbon

[0]Inst. J. Monod

[0]Weizmann Institute

4.5. Protein polymerisation

Protein polymerisation is the key feature of amyloid diseases, among which we can quote Alzheimer's, Prion (in particular variant Creutzfeldt-Jakob disease, epidemically linked to bovine spongiform encephalopathy, or so-called "mad cow", disease), Parkinson's, Huntington's diseases. However, the mechanisms of polymerisation are far from being quantitatively understood by biologists. They can be modelled with the help of coagulation-fragmentation equations, a field of expertise of MAMBA [51], [50]. Hence the European starting grant SKIPPER^{AD}, which follows the ANR project TOPPAZ, came up very naturally from the encounter with Human Rezaei, a biologist expert in amyloid diseases at INRA Jouy-en-Josas. Moreover, this field of applications brings new questions to us, which is both a stimulation for our mathematical research and a very promising tool for the biologists.

4.6. Physics of tissue organisation

Many new insights in the last years indicate that migration, growth and division of cells are largely impacted by cell and tissue mechanics ([0], [0], [0]). Centre-based growth models already permit to account for many of the observed phenomena (e.g. [0], [0]). They furthermore permit calculation of the stress tensor in the tissue. ABMs resolving cells at higher resolution [0] permit to calculate cell deformation as a response of stress emerging in the tissue, hence the stress tensor cannot only be resolved at the position of the cell centre, as in the case of centre-based models, but in this case at any point on the cell surface or inside the cell. This permits relating stress and strain in tissues and the deformation and stress a cell feels at subcellular scale. We extended a deformable cell model towards cell-division which enables us to calculate precise stress - strain relationships for cells, that later can be used to calibrate forces in center-based models. This is fundamental to understand the impact of mechanical stress on cell cycle progression or other cell decisions. Moreover, we established a model to explain the proliferation pattern of cells growing in closed capsules.

4.7. Liver modelling

Liver is the main detoxifying organ of the human body and can regenerate up to about 70% of its mass. It pursues its task due to a complex tissue architecture, with hepatocytes aligning along micro-capillaries and forming a dense network. Incidence rate of liver diseases are steadily increasing, liver cancer ranks 6th among all cancers. About one person in 12 suffer from viral hepatitis, which makes 500 million people worldwide. Hepatitis B and C as well as misuse of drugs or alcohol are major reasons to develop liver cancer. Notwithstanding the importance of this public health problem, disease pathogenesis and regeneration in liver are still not well understood.

So far systems biology approaches addressing the tissue scale are rare. Most of those which do so base on compartment models (e.g. [0]); only recently are being developed approaches addressing the tissue scale ([57], [0], [0], [0], [0]). We are developing a multi-scale model of liver regeneration representing the tissue architecture, the different cell types, the flow systems, hepatocyte metabolism and signal transduction controlling cell cycle entrance in the regeneration processes, taking into account extrahepatic compartments when relevant. Applications are regeneration after drug-induced damage and after partial hepatectomy, drug pharmacodynamics and pharmacokinetics in liver and liver cancer, and model-based prediction of in-vivo drug toxicity from in-vitro measurements [0]. The research work is performed within the EU project NOTOX, the BMBF project VIRTUAL LIVER NETWORK and the ANR project IFLOW.

---

[0]Ingber, Proc. Natl. Acad. Sci (USA), 2005
[0]Trepat et. al., Nat. Phys. 2009
[0]Alessandri et. al., Proc. Natl. Acad. Sci. (USA) 2013
[0]Drasdo and Hoehme, Phys. Biol. 2005
[0]Drasdo and Hoehme, New Journal of Physics 2012
[0]Odenthal, Smeets, van Liedekerke, et. al., PloS Comput Biol. 2013
[0]Diaz-Ochoa et. al. Frontiers in Pharmacology, 2013
[0]Ricken, Dahmen, Dirsch, Biomech. Model. Mechanobiol. 2010
[0]Debbaut et. al., J. Biomech. Eng. 2014
[0]Siggers, Leungchavphongse, Ho, Repetto, Biomech. Model. Mechanobiol. 2014
[0]Schwen et. al., PloS Comput. Biol. 2014
[0]Godoy et al., *Arch Toxicol. 2013* Aug;87(8):1315-1530

MYCENAE Project-Team

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.

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 closed 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 [5], [6], analysis of neurosecretory patterns [13] and modeling of GPCR (G-Protein Coupled Receptors) signaling networks [8]. 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 [42].

POMDAPI Project-Team

3. Application Domains

3.1. Environmental sciences

Applications are in hydrogeology and water resources.

3.2. Energy sciences

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

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

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.* [62], [66], [65] 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.

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

3. Application Domains

3.1. Medical applications of modeling, signal analysis and control

- *3D cardiac modeling for personalized medicine.* Our main contribution to Inria's collective efforts in this field (project-teams Asclepios, MACS, REO, Sisyphe) is the so-called "Bestel-Clément-Sorine" model of contraction of cardiac muscle [39], at the origin of the 3D electromechanical direct and inverse modeling of the heart at Inria. This model is based on ideas originating from the kinetic equation theory, used to model, on the molecular scale, the controlled collective behavior of actin-myosin nanomotors at the root of muscle contraction. The classical Huxley model was recovered on the sarcomere scale by using moment equations and a controlled constitutive law on the tissue scale was obtained using the same type of scaling techniques. The model, now embedded in heart simulators, is used in various studies [51], [3], [54], [53].

- *Semiclassical analysis of cardiovascular signals.* This work began with the article [41] and the PhD of M. Laleg-Kirati [47], [46], [49]. The theory and a validation of a new method of blood pressure analysis are now published [17], [48].
The main idea is to consider a signal $x \to y(x)$ to analyze as the multiplication operator $\phi \to y\phi$ on some function space, and to analyze it as a potential. The signal is represented by the spectrum of an associated Schrödinger operator, combined with a semi-classical quantification: $-h^2 \dfrac{d^2}{dx^2} - y(x)$ with $h > 0$ small. For signals looking as "superpositions of bumps" (e.g. the systolic pulse, the dichrotic notch for the arterial pulse pressure), this leads to some kind of nonlinear Fourier analysis [17]. The spectral parameters associated with the arterial pressure can be useful cardiovascular indices, e.g. for noninvasive blood flow estimation [48]. In the arterial pressure case, this is equivalent to approximate the traveling pressure pulse by a N-soliton solution of a Korteweg-de Vries (KdV) equation [41] and using ideas similar to the Lax pair representation of N-solitons and proof technique for the weak dispersion limit of KdV. A striking result is that an N-soliton is a very good representation of the arterial pressure waveform for values of N as small as $N = 3$. The representation of pulse-shaped signals is parsimonious, having only $2N$ parameters [55].

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

- *Modeling and control of CARMAT Total Artificial Heart.* This TAH has been implanted for the first time in a patient in Dec 2013. We have contributed to this industrial project since 2008 on modeling and control questions during the post-doc of Karima Djabella (now engineer at CARMAT), Frédéric Vallais and the two-year contract for supervising Julien Bernard (CARMAT control engineer). It was an opportunity to exploit some results on the baroreflex control [42] or heart rate variability during exercise [40].

- *Glycemic control in Intensive Care Units (ICUs):* Blood glucose is a key biological parameter in ICU since the study of van den Berghe et al [60] who demonstrated decreased mortality in surgical intensive care patients in association with tight glycemic control (TGC), based on intensive insulin therapy. In their work, however, there was only one ICU and the protocol was not formalized. Trying to decrease mortality in standard ICUs by using computer aided glycemic control is still a challenge. Previous studies have failed because of high rates of severe hypoglycemia. The last one was NICE-SUGAR [57] with a 2% increase in mortality (death ratio from any cause within 90 days after randomization compared between control and TGC patients). In cooperation with Pierre Kalfon (Intensive Care, Hospital of Chartres) and in the framework of a CIFRE contract with a small medtech company LK2 (Tours, France), we have studied the origins of these failures and proposed more robust control algorithms tuned using a database of representative "virtual patients". See [44] and the PhD of A. Guerrini, [43]. A first version of the controller has been tested in a large clinical study CGAO-REA [14].

3.2. Engineering applications of modeling, signal analysis and control

Identification of nonlinear systems: from algorithms to a popular matlab toolbox:
Block-oriented nonlinear system identification with Jiandong Wang (Associate Professor, Beijing University, China) [58]; Development of the Matlab System Identification ToolBox (SITB).

Identification of transmission line characteristics: from algorithms to electronic experiments. Collaboration with CEA LIST (Lab of applied research on software-intensive technologies) and LGEP (Laboratoire de génie électrique de Paris) with Florent Loete [50] (ANR projects SEEDS, 0-DEFECT, INSCAN, SODDA).
We have extended to some networks the seminal work of Jaulent [45] for the real line: all the information contained in a measured reflection coefficient can be obtained by solving an inverse scattering problem for a system of Schrödinger or Zakharov-Shabat equations on the graph of the network, which allows one to recover the geometry of the network and some electrical characteristics for nonuniform lossless electrical star-shaped networks [26]. An efficient method to solve the associated Guelfand-Levitan-Marchenko equations has been studied and is used in the software ISTL that has been developed in Sisyphe [59], [56]. This development will continue in the project-team I4S. An engineering methodology based on this approach has been described [29] and some first experimental results obtained [36], [50].

Monitoring and control of automotive depollution systems: with RENAULT (Karim Bencherif, Damiano Di Penta and PhD students): [52], [20], [38].

3.3. Modeling and optimizing patient pathways in hospital

This research theme was inspired by the observation that the practice of modern industrialized medical care proceeds by queueing and forwarding patients from one step of care to the next, with each step involving specialized personnel and machinery. Whereas the human competence and machine performance available in hospital are both highly evolved, the deployment of these resources for the patient's benefit is problematic.

If we change our point of view and, rather than look in the traditional way at the hospital as a collection of "vertical" silos (emergency room, cardiac ward, respiratory ward, operating rooms, imagery services, blood services, logistics, etc.), we instead consider the patient's "horizontal" trajectory crossing through many different silos, we can see significant problems. In between each step, whose value added part typically lasts at most a few minutes, there are long stretches of time (on the order of hours) during which the patient simply waits for the hospital to arrange the next step of care — with serious consequences:
- delayed treatment is directly correlated with increased rates of hospital acquired nosocomial infection for the immuno-compromised and to loss of autonomy for the elderly;
- the cost in terms of scarce resources (for example, bed-hours in hospital) is significant, but the root causes of these costs due to problems at the interfaces between silos are hard to observe since they do not fit neatly into the hospital's traditional hierarchy;
- the slow and error prone hand-offs of information between silos are dangerous for the patient who is vulnerable to medical errors. In the US, for example, it is estimated that there are 100,000 deaths per year attributable to hospital error (triple the number due to road traffic accidents by way of comparison).

<div align="center">

ALPINES Project-Team

</div>

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 sub-region of the propagation medium. Although a huge literature is already available for the study of localized singular perturbations and boundary layer pheneomena, very few works have proposed efficient numerical methods that rely on asymptotic modeling. This is due to their natural functional framework that naturally involves singular functions, which are difficult handle numerically. The aim of this part of our reasearch is to develop and analyze numerical methods for singular perturbation methods that are prone to high order numerical approximation, and robust with respect to the small parameter caracterizing the singular perturbation.

4.4. Data analysis in astrophysics

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

In CMB data analysis, astrophysicists produce and analyze multi-frequency 2D images of the universe when it was 5% of its current age. The new generation of the CMB experiments observes the sky with thousands of detectors over many years, producing overwhelmingly large and complex data sets, which nearly double every year therefore following the Moore's Law. Planck (http://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.

DYOGENE Project-Team

4. Application Domains

4.1. Embedded Networks

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

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

4.2. Wireless Networks

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

4.3. Peer-to-Peer Systems

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

4.4. Social and Economic Networks

Networks are ubiquitous with the presence of different kinds of social, economic and information networks around us. The Internet is one of the most prominent examples of a geometric network. We also examine geometric networks from the perspective of sociologist and economist [55]. Network analysis is also attracting fundamental research by computer scientists [49]. Diffusion of information, social influence, trust, communication and cooperation between agents are heavily researched topics in e-commerce and multi-agent systems. Our probabilistic techniques are very appropriate in this case and have been largely neglected so far. While the first works on geometric networks emanated from theoretical physicists, they stay more focused on static properties of such networks and do not consider game theoretical or statistical learning (like community detection) aspects of such networks. This leaves open a range of new problems to which we will contribute.

4.5. Routing Protocols

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

GANG Project-Team

4. Application Domains

4.1. Application Domains

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

- The application of measuring and modeling Internet metrics such as latencies and bandwidth is to provide tools for optimizing Internet applications. This concerns especially large scale applications such as web site mirroring and peer-to-peer applications.

- Peer-to-peer protocols are based on a all equal paradigm that allows to design highly reliable and scalable applications. Besides the file sharing application, peer-to-peer solutions could take over in web content dissemination resistant to high demand bursts or in mobility management. Envisioned peer-to-peer applications include video on demand, streaming, exchange of classified ads,...

- Wifi networks have entered our every day life. However, enabling them at large scale is still a challenge. Algorithmic breakthrough in large ad hoc networks would allow to use them in fast and economic deployment of new radio communication systems.

- The main application of the web graph structure consists in ranking pages. Enabling site level indexing and ranking is a possible application of such studies.

HIPERCOM2 Team

4. Application Domains

4.1. Introduction

The HIPERCOM2 team addresses the following application domains:

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

These application domains use the four types of wireless networks:

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

4.2. Wireless mesh and mobile ad hoc networks

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

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

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

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

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

4.3. Vehicular Networks and Smart Cities

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

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

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

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

- Energy efficiency is a key property in wireless sensor networks. Various techniques contribute to save energy of battery-equipped sensor nodes. To name a few, they are: energy efficient routing protocols, node activity scheduling, adjustment of transmision power, reduction of protocols overhead, reduction of data generated and transmitted. In the OCARI network, an industrial wireless sensor network, we have designed and implemented an energy efficient routing protocol and a node activity scheduling algorithm allowing router nodes to sleep. We have applied a cross-layering approach allowing the optimization of MAC and network protocols taking into account the application requirements and the environment in which the network operates. We have observed the great benefit obtained with node activity scheduling. In networks with low activity, opportunistic strategies are used to address low duty cycles.

- Large scale WSNs constitute another challenge. Large autonomous wireless sensors in the internet of the things need very well tuned algorithms. Self-organization is considered as a key element in tomorrow's Internet architecture. A major challenge concerning the integration of self-organized networks in the Internet is the accomplishment of light weight network protocols in large ad hoc environments.

- Multichannel WSNs provide an opportunity:
    - to increase the parallelism between transmissions. Hence, it reduces the data gathering delays and improves the time consitency of gathered data.
    - to increase the robustness against interferences and perturbations possibly caused by the coexistence of other wireless networks.

4.5. Cognitive Radio Networks

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

- Design and modeling of a new access scheme based on a generalized Carrier Sense Multiple Access scheme using active signaling. This scheme allows the primary users to capture the bandwidth even if the secondary users are transmitting in the area.

- Design of a time slot and channel assignment to minimize the data gathering performed by secondary users. This assignment should work with different detection schemes of primary user presence.

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

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 have 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* under creation:

– **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 the effort of creating CityLab@Inria [0]. 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, OAK, SMIS, URBANET and WILLOW. CityLab further involves strong collaboration with Californian 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 on leave 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).

---

[0] https://citylab.inria.fr

MUSE Team

4. Application Domains

4.1. 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 "Internet 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 heterogenous. 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 *WTF*) 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 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.

4.2. Quality of Experience

Understanding how users react to different levels of network performance presents two main challenges:

1. User perception is subjective and contextual. Different users may have different tolerance levels to network performance and the same user may have different expectations under different circumstances. Take for example the round-trip time (RTT), a typical network performance metric. If RTTs are larger than usual, a user who is doing remote login may feel that the connection is unusable, whereas another who is watching YouTube may notice no problem (because YouTube has a playout buffer to mask some network delay). Take another example of a user downloading her email. This user may tolerate some delay when she is leisurely checking her email at home, but she may become extremely frustrated with the same delay if she is in an airplane and needs to download her email just before takeoff.

2. It is challenging to "measure" users. We must develop methods to measure the user perception of network performance as users perform their routine online tasks. It is hence important that these methods are not too intrusive. Otherwise, users are unlikely to participate in the experiment. In addition, we must capture user perception at different levels of performance and in a variety of scenarios.

We will develop tools that run on end systems to collect network performance data annotated with the user perception. These tools will adopt a hybrid measurement methodology that combines network measurement techniques to infer application performance with techniques from HCI to measure user perception. We will later use the resulting datasets to build models of user perception of network performance based only on data that we can obtain automatically from the user device or from user's traffic observed in the network. Models of user perception of network performance can be used to detect when performance is poor to trigger diagnosis or to adapt network/application performance to better serve users.

4.3. Crowd-sourced content recommendation

The Internet today serves as a large content distribution platform (online content varies from traditional news, TV series, and movies to specialized blogs and family pictures shared over social networks) as well as a platform for users to exchange opinions about practically everything (from movies to services and restaurants). The amount of information available online today overwhelms most users and selecting which content to watch or what do has become a challenge. We are applying passive measurement methods and content summarisation techniques to help users to identify relevant content in two scenarios. First, we are developing a system called WeBrowse that passively observes network traffic to extract user clicks (i.e., the URLs users visit). A user click is a good measure of interest, as users often have an idea of the type of content they are about to access (e.g., because they saw a preview or because a friend recommended it). Intuitively, the more users click on a URL, the higher the interest in the content on the corresponding page. WeBrowse then promotes "hottest" and most popular content to users of a network. We have a deployment of WeBrowse in a campus network. Second, we are working on techniques to summarise user feedback (for example, movie or restaurant reviews) with semi-structured feedback. Today reviews are either free-form text or star rating. Star rating is too coarse to capture the nuances of why a user likes or dislikes something, whereas free text is hard for users to parse and

extract a clear opinion. We are instead working with semi-structured reviewing where users enter *tags* (a short sequence of words describing the user experience). We are working with Technicolor on the summarisation of movie reviews and on building a mobile app (called TagIt) where users can review movies directly with tags.

RAP Project-Team (section vide)

REGAL Project-Team (section vide)

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, v3.17, comprises over 12 million lines of code, and supports 29 different families of CPU architectures, 73 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. [42], numerous research tools have been applied to the Linux kernel, typically for finding bugs [41], [56], [65], [72] or for computing software metrics [46], [74]. In our work, we have studied generic C bugs in Linux code , bugs in function protocol usage [50], [51], issues related to the processing of bug reports [20] and crash dumps [18], and the problem of backporting (work in progress), 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. Section 5.1.2 presents our dissemination efforts to the Linux community.

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, [36] a template-based approach for porting device drivers from one OS to another. Ryzhyk *et al.* propose Termite, [66], [67] 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 [47] 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, [27] Coverity [42], CP-Miner, [55] PR-Miner [56], 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 guidelines and tools for the maintenance and the development of new drivers. Section 3.2 describes this research direction.

ALPAGE Project-Team

4. Application Domains

4.1. Overview

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

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

Language writing aid : spelling, grammatical and stylistic correctors for text editors, controlled-language writing aids (e.g., for technical documents), memory-based translation aid, foreign language learning tools, as well as vocal dictation; related to this group lies the automatic correction of the output of OCR systems;

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

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

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

4.2. Information extraction and knowledge acquisition

Participants: Éric Villemonte de La Clergerie, Benoît Sagot.

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

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

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

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

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

Participants: Benoît Sagot, Valérie Hanoka.

Verbatim Analysis is a startup co-created by Benoît Sagot from Alpage and Dimitri Tcherniak from Towers Watson, a world-wide leader in the domain of employee research (opinion mining among the employees of a company or organization). The aim of its first product, *vera*, is to provide an all-in-one environment for editing (i.e., normalizing the spelling and typography), understanding and classifying answers to open-ended questions, and relating them with closed-ended questions, so as to extract as much valuable information as possible from both types of questions. The editing part relies in part on SxPipe (see section 5.5) and Alexina morphological lexicons. Several other parts of *vera* have been co-developed by Verbatim Analysis and by Inria.

4.4. Multilingual terminologies and lexical resources for companies

Participant: Éric Villemonte de La Clergerie.

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

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

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

NLP tools and resources used for spelling correction, such as large n-gram collections, POS taggers and finite-state machinery are now mature and precise. In industrial setting such as post-processing after large-scale OCR, these tools and resources should enable spelling correction tools to work on a much larger scale and with a much better precision than what can be found in different contexts with different constraints (e.g., in text editors). Moreover, such industrial contexts allow for a non-costly manual intervention, in case one is able to identify the most uncertain corrections. Alpage is working within the "Investissements d'avenir" project PACTE, headed by Numen, a company specialized in text digitalization, and three other partners. Kata Gábor and Pierre Magistry are doing post-docs funded by PACTE (see 6.3)

4.6. Experimental and quantitative linguistics

Participants: Benoit Crabbé, Benoît Sagot, Alexandra Simonenko, Sarah Beniamine, Kristina Gulordava.

Alpage is a team that dedicates efforts in producing ressources and algorithms for processing large amounts of textual materials. These ressources can be applied not only for purely NLP purposes but also for linguistic purposes. Indeed, the specific needs of NLP applications led to the development of electronic linguistic resources (in particular lexica, annotated corpora, and treebanks) that are sufficiently large for carrying statistical analysis on linguistic issues. In the last 10 years, pioneering work has started to use these new data sources to the study of English grammar, leading to important new results in such areas as the study of syntactic preferences [62], [139], the existence of graded grammaticality judgments [86].

The reasons for getting interested for statistical modelling of language can be traced back by looking at the recent history of grammatical works in linguistics. In the 1980s and 1990s, theoretical grammarians have been mostly concerned with improving the conceptual underpinnings of their respective subfields, in particular through the construction and refinement of formal models. In syntax, the relative consensus on a generative-transformational approach [72] gave way on the one hand to more abstract characterizations of the language faculty [72], and on the other hand to the construction of detailed, formally explicit, and often implemented, alternative formulation of the generative approach [61], [103]. For French several grammars have been implemented in this trend, such as the tree adjoining grammars of [65], [76] among others. This general movement led to much improved descriptions and understanding of the conceptual underpinnings of both linguistic competence and language use. It was in large part catalyzed by a convergence of interests of logical, linguistic and computational approaches to grammatical phenomena.

However, starting in the 1990s, a growing portion of the community started being frustrated by the paucity and unreliability of the empirical evidence underlying their research. In syntax, data was generally collected impressionistically, either as ad-hoc small samples of language use, or as ill-understood and little-controlled grammaticality judgements [121]. This shift towards quantitative methods is also a shift towards new scientific questions and new scientific fields. Using richly annotated data and statistical modelling, we address questions that could not be addressed by previous methodology in linguistics.

In this line, at Alpage we have started investigating the question of choice in French syntax with a statistical modelling methodology. In the perspective of better understanding which factors influence the relative ordering of post verbal complements across languages and through language evolution.

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

In parallel, quantitative methods are applied to computational morphology, in particlar in the context of Sarah Beniamine's PhD, co-supervized by Benoît Sagot (Alpage) and Olivier Bonami (LLF, CNRS, U. Paris Diderot and U. Paris Sorbonne). Collaborative work in this area is also conducted in collaboration with descriptive linguists from CRLAO (CNRS and Inalco; Guillaume Jacques) and HTL (CNRS, U. Paris Diderot and U. Sorbonne Nouvelle; Aimée Lahaussois) and formal linguists from DDL (CNRS and Université Lyon 2; Géraldine Walther).

<p style="text-align:center">**RITS 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, new transportation systems 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 into account.

4.3. New transportation systems

The problems related to the abusive use of the individual car in large cities led the populations and the political leaders to support the development of public transport. A demand exists for a transport of people and goods which associates quality of service, environmental protection and access to the greatest number. Thus the Tram and the light subways of VAL type recently introduced into several cities in France conquered the populations, in spite of high financial costs. However, these means of mass transportation are only possible on lines on which there is a keen demand. As soon as one moves away from these "lines of desire" or when one deviates from the rush hours, these modes become expensive and offer can thus only be limited in space and time. To give a more flexible offer, it is necessary to plan more individual modes which approach the car as we know it. However, if one wants to enjoy the benefits of the individual car without suffering from their disadvantages, it is necessary to try to match several criteria: availability anywhere and anytime to all, lower air and soils pollution as well as sound levels, reduced ground space occupation, security, low cost. Electric or gas vehicles available in self-service, as in the Praxitèle system, bring a first response to these criteria. To be able to still better meet the needs, it is however necessary to re-examine the design of the vehicles on the following points:

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

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

4.4. Automated vehicles

The long term effort of the project is to put automatically guided vehicles (cybercars) on the road. It seems too early to mix cybercars and traditional vehicles, but data processing and automation now make it possible to consider in the relatively short term the development of such vehicles and the adapted infrastructures. 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 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. All our developments and demonstrations on our cybercars (cycabs, Yamaha AGV and new Cybus platforms) are using the RTMaps SDK development platform. These demonstrations include: reliable SLAMMOT algorithm using 2 to 4 laser sensors simultaneously, automatic line/road following techniques, PDA remote control, multi sensors data fusion, collaborative perception via ad-hoc network. The second main topic is inter-vehicle communications using ad-hoc networks. We have worked with the HIPERCOM team for setting and tuning OLSR, a dynamic routing protocol for vehicles communications. 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.

4. Application Domains

4.1. Application Domains

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

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

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

WILLOW Project-Team

4. Application Domains

4.1. Introduction

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

4.2. Quantitative image analysis in science and humanities

We plan to apply our 3D object and scene modeling and analysis technology to image-based modeling of human skeletons and artifacts in anthropology, and large-scale site indexing, modeling, and retrieval in archaeology and cultural heritage preservation. Most existing work in this domain concentrates on image-based rendering—that is, the synthesis of good-looking pictures of artifacts and digs. We plan to focus instead on quantitative applications. We are engaged in a project involving the archaeology laboratory at ENS and focusing on image-based artifact modeling and decorative pattern retrieval in Pompeii. 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.