Activity Report 2012

Section Scientific Foundations

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

3. Scientific Foundations

3.1. Models of Computation and Communication (MoCCs)

Participants: Charles André, Robert de Simone, Jean-Vivien Millo, Dumitru Potop Butucaru.

Esterel, SyncCharts, synchronous formalisms, Process Networks, Marked Graphs, Kahn networks, compilation, synthesis, formal verification, optimization, allocation, refinement, scheduling

Formal Models of Computation form the basis of our approach to Embedded System Design. Because of the growing importance of communication handling, it is now associated with the name, MoCC in short. The appeal of MoCCs comes from the fact that they combine features of mathematical models (formal analysis, transformation, and verification) with this of executable specifications (close to code level, simulation, and implementation). Examples of MoCCs in our case are mainly synchronous reactive formalisms and dataflow process networks. Various extensions or specific restrictions enforce respectively greater expressivity or more focused decidable analysis results.

DataFlow Process Networks and Synchronous Reactive Languages such as ESTEREL/SYNCCHARTS and SIGNAL/POLYCHRONY [53], [54], [48], [15], [4], [13] share one main characteristics: they are specified in a self-timed or loosely timed fashion, in the asynchronous data-flow style. But formal criteria in their semantics ensure that, under good correctness conditions, a sound synchronous interpretation can be provided, in which all treatments (computations, signaling communications) are precisely temporally mapped. This is refered to as clock calculus in synchronous reactive systems, and leads to a large body of theoretical studies and deep results in the case of DataFlow Process Networks [49], [47] (consider SDF balance equations for instance [56]).

As a result, explicit schedules become an important ingredient of design, which ultimately can be considered and handled by the designer him/herself. In practice such schedules are sought to optimize other parts of the design, mainly buffering queues: production and consumption of data can be regulated in their relative speeds. This was specially taken into account in the recent theories of Latency-Insensitive Design [50], or N-synchronous processes [51], with some of our contributions [6].

Explicit schedule patterns should be pictured in the framework of low-power distributed mapping of embedded applications onto manycore architectures, where they could play an important role as theoretical formal models on which to compute and optimize allocations and performances. We describe below two lines of research in this direction. Striking in these techniques is the fact that they include time and timing as integral parts of early functional design. But this original time is logical, multiform, and only partially ordering the various functional computations and communications. This approach was radically generalized in our team to a methodology for logical time based design, described next (see 3.2).

3.1.1. *K-periodic static scheduling and routing in Process Networks*

In the recent years we focused on the algorithm treatments of ultimately k-periodic schedule regimes, which are the class of schedules obtained by many of the theories described above. An important breakthrough occurred when realizing that the type of ultimatelly periodic binary words that were used for reporting *static scheduling* results could also be employed to record a completely distinct notion of ultimately k-periodic route switching patterns, and furthermore that commonalities of representation could ease combine them together. A new model, by the name of K-periodical Routed marked Graphs (KRG) was introduced, and extensively studied for algebraic and algorithmic properties [5].

The computations of optimized static schedules and other optimal buffering configurations in the context of latency-insensitive design led to the K-Passa software tool development 5.2 .

3.1.2. Endochrony and GALS implementation of conflict-free polychronous programs

The possibility of exploring various schedulings for a given application comes from the fact that some behaviors are truly concurrent, and mutually *conflict-free* (so they can be executed independently, with any choice of ordering). Discovering potential asynchronous inside synchronous reactive specifications then becomes something highly desirable. It can benefit to potential distributed implementation, where signal communications are restricted to a minimum, as they usually incur loss in performance and higher power consumption. This general line of research has come to be known as Endochrony, with some of our contributions [11].

3.2. Logical Time in Model-Driven Embedded System Design

Participants: Charles André, Julien deAntoni, Frédéric Mallet, Marie-Agnès Peraldi Frati, Robert de Simone.

Starting from specific needs and opportunities for formal design of embedded systems as learned from our work on MoCCs (see 3.1), we developed a Logical Time Model as part of the official OMG UML profile MARTE for Modeling and Analysis of Real-Time Embedded systems. With this model is associated a Clock Constraint Specification Language (CCSL), which allows to provide loose or strict logical time constraints between design ingredients, be them computations, communications, or any kind of events whose repetitions can be conceived as generating a logical conceptual clock (or activation condition). The definition of CCSL is provided in [1].

Our vision is that many (if not all) of the timing constraints generally expressed as physical prescriptions in real-time embedded design (such as periodicity, sporadicity) could be expressed in a logical setting, while actually many physical timing values are still unknown or unspecified at this stage. On the other hand, our logical view may express much more, such as loosely stated timing relations based on partial orderings or partial constraints.

So far we have used CCSL to express important phenonema as present in several formalisms: AADL (used in avionics domain), EAST-ADL2 (proposed for the AutoSar automotive electronic design approach), IP-Xact (for System-on-Chip (*SoC*) design). The difference here comes from the fact that these formalisms were formerly describing such issues in informal terms, while CCSL provides a dedicated formal mathematical notation. Close connections with synchronous and polychronous languages, especially Signal, were also established; so was the ability of CCSL to model dataflow process network static scheduling.

In principle the MARTE profile and its Logical Time Model can be used with any UML editor supporting profiles. In practice we focused on the PAPYRUS open-source editor, mainly from CEA LIST. We developed under Eclipse the TIMESQUARE solver and emulator for CCSL constraints (see 5.1), with its own graphical interface, as a stand-alone software module, while strongly coupled with MARTE and Papyrus.

While CCSL constraints may be introduced as part of the intended functionality, some may also be extracted from requirements imposed either from real-time user demands, or from the resource limitations and features from the intended execution platform. Sophisticated detailed descriptions of platform architectures are allowed using MARTE, as well as formal allocations of application operations (computations and communications) onto platform resources (processors and interconnects). This is of course of great value at a time where embedded architectures are becoming more and more heterogeneous and parallel or distributed, so that application mapping in terms of spatial allocation and temporal scheduling becomes harder and harder. This approach is extensively supported by the MARTE profile and its various models. As such it originates from the Application-Architecture-Adequation (AAA) methodology, first proposed by Yves Sorel, member of Aoste. AAA aims at specific distributed real-time algorithmic methods, described next in 3.3 .

Of course, while logical time in design is promoted here, and our works show how many current notions used in real-time and embedded systems synthesis can naturally be phrased in this model, there will be in the end a phase of validation of the logical time assumptions (as is the case in synchronous circuits and SoC design with timing closure issues). This validation is usually conducted from Worst-Case Execution Time (WCET) analysis on individual components, which are then used in further analysis techniques to establish the validity of logical time assumptions (as partial constraints) asserted during the design.

3.3. The AAA (Algorithm-Architecture Adequation) methodology and Real-Time Scheduling

Participants: Laurent George, Dumitru Potop Butucaru, Yves Sorel.

Note: The AAA methodology and the SynDEx environment are fully described at http://www.syndex.org/, together with relevant publications.

3.3.1. *Algorithm-Architecture Adequation*

The AAA methodology relies on distributed real-time scheduling and relevant optimization to connect an Algorithm/Application model to an Architectural one. We now describe its premises and benefits.

The Algorithm model is an extension of the well known data-flow model from Dennis [52]. It is a directed acyclic hyper-graph (DAG) that we call "conditioned factorized data dependence graph", whose vertices are "operations" and hyper-edges are directed "data or control dependences" between operations. The data dependences define a partial order on the operations execution. The basic data-flow model was extended in three directions: first infinite (resp. finite) repetition of a sub-graph pattern in order to specify the reactive aspect of real-time systems (resp. in order to specify the finite repetition of a sub-graph consuming different data similar to a loop in imperative languages), second "state" when data dependences are necessary between different infinite repetitions of the sub-graph pattern introducing cycles which must be avoided by introducing specific vertices called "delays" (similar to z^{-n} in automatic control), third "conditioning" of an operation by a control dependence similar to conditional control structure in imperative languages, allowing the execution of alternative subgraphs. Delays combined with conditioning allow the programmer to specify automata necessary for describing "mode changes".

The Architecture model is a directed graph, whose vertices are of two types: "processor" (one sequencer of operations and possibly several sequencers of communications) and "medium" (support of communications), and whose edges are directed connections.

The resulting implementation model [9] is obtained by an external compositional law, for which the architecture graph operates on the algorithm graph. Thus, that result is a set of algorithm graphs, "architecture-aware", corresponding to refinements of the initial algorithm graph, by computing spatial (distribution) and timing (scheduling) allocations of the operations according to the architecture graph resource availability. In that context "Adequation" refers to some search amongst the solution space of resulting algorithm graphs, labelled by timing characteristics, for one which verifies timing constraints and optimizes some criteria, usually the total execution time and the number of computing resources (but other criteria may exist). The next section describes distributed real-time schedulability analysis and optimization techniques for that purpose.

3.3.2. *Distributed Real-Time Scheduling and Optimization*

We address two main issues: monoprocessor real-time scheduling and multiprocessor real-time scheduling where constraints must mandatorily be met otherwise dramatic consequences may occur (hard real-time) and where resources must be minimized because of embedded features.

In our monoprocessor real-time scheduling work, beside the classical deadline constraint, often equal to a period, we take into consideration dependences beetween tasks and several, possibly related, latencies. A latency is a generalization of the typical "end-to-end" constraint. Dealing with multiple real-time constraints raises the complexity of that issue. Moreover, because the preemption leads to a waste of resources due to its approximation in the WCET (Worst Execution Time) of every task as proposed by Liu and Leyland [57], we first studied non-preemtive real-time scheduling with dependences, periodicities, and latencies constraints. Although a bad approximation may have dramatic consequences on real-time scheduling, there are only few researches on this topic. We have been investigating preemptive real-time scheduling since few years, but seeking the exact cost of the preemption such that it can be integrated in schedulability conditions, and in the corresponding scheduling algorithms. More generally, we are interested in integrating in the schedulability analyses the cost of the RTOS (Real-Time Operating System), for which the exact cost of preemption is the most difficult part because it varies according to the instance of each task [10]. Finally, we investigate also the problem of mixing hard real-time and soft real-time constraints that arises in the most complex applications.

The second research area is devoted to distributed real-time scheduling with embedding constraints. We use the results obtained in the monoprocessor case in order to derive solutions for the problem of multiprocessor (distributed) real-time scheduling. In addition to satisfy the multiple real-time constraints mentioned in the monoprocessor case, we have to minimize the total execution time (makespan) since we deal with automatic control applications involving feedback. Furthermore, the domain of embedded systems leads to solving minimization resources problems. Since these optimization problems are of NP-hard complexity we develop exact algorithms (B & B, B & C) which are optimal for simple problems, and heuristics which are sub-optimal for realistic problems corresponding to industrial needs. Long time ago we proposed a very fast "greedy" heuristics [8] whose results were regularly improved, and extended with local neighborhood heuristics, or used as initial solutions for metaheuristics such as variants of "simulated annealing".

In addition to the spatial dimension (distributed) of the real-time scheduling problem, other important dimensions are the type of communication mechanisms (shared memory vs. message passing), or the source of control and synchronization (event-driven vs. time-triggered). We explore real-time scheduling on architectures corresponding to all combinations of the above dimensions. This is of particular impact in application domains such as automotive and avionics (see 4.2).

Since real-time distributed systems are often safety-critical we address dependability issues, to tolerate faults in processors and communication interconnects. We maily focus on software redondancy, rather than hardware, to ensure real-time behaviour preservation in presence of faulty processors and/or communication media (where possible failures are predictively specified by the designer). We investigate fail silent, transient, intermittent, and Byzantine faults.

GALAAD Project-Team

3. Scientific Foundations

3.1. Introduction

Our scientific activity is structured according to three broad topics:

1. **Algebraic representations for geometric modeling**.
2. **Algebraic algorithms for geometric computing**,
3. **Symbolic-numeric methods for analysis**,

3.2. Algebraic representations for geometric modeling

Compact, efficient and structured descriptions of shapes are required in many scientific computations in engineering, such as "Isogeometric" Finite Elements methods, point cloud fitting problems or implicit surfaces defined by convolution. Our objective is to investigate new algebraic representations (or improve the existing ones) together with their analysis and implementations.

We are investigating representations, based on semi-algebraic models. Such non-linear models are able to capture efficiently complexes shapes, using few data. However, they required specific methods to solve the underlying non-linear problems, which we are investigating.

Effective algebraic geometry is a natural framework for handling shape representations. This framework not only provides tools for modeling but it also allows to exploit rich geometric properties.

The above-mentioned tools of effective algebraic geometry make it possible to analyse in detail and separately algebraic varieties. We are interested in problems where collections of piecewise algebraic objects are involved. The properties of such geometrical structures are still not well controlled, and the traditional algorithmic geometry methods do not always extend to this context, which requires new investigations.

The use of piecewise algebraic representations also raises problems of approximation and reconstruction, on which we are working on. In this direction, we are studying B-spline function spaces with specified regularity associated to domain partitions.

Many geometric properties are, by nature, independent from the reference one chooses for performing analytic computations. This leads naturally to invariant theory. We are interested in exploiting these invariant properties, to develop compact and adapted representations of shapes.

3.3. Algebraic algorithms for geometric computing

This topic is directly related to polynomial system solving and effective algebraic geometry. It is our core expertise and many of our works are contributing to this area.

Our goal is to develop algebraic algorithms to efficiently perform geometric operations such as computing the intersection or self-intersection locus of algebraic surface patches, offsets, envelopes of surfaces, ...

The underlying representations behind the geometric models we consider are often of algebraic type. Computing with such models raises algebraic questions, which frequently appear as bottlenecks of the geometric problems.

In order to compute the solutions of a system of polynomial equations in several variables, we analyse and take advantage of the structure of the quotient ring, defined by these polynomials. This raises questions of representing and computing normal forms in such quotient structures. The numerical and algebraic computations in this context lead us to study new approaches of normal form computations, generalizing the well-known Gröbner bases.

Geometric objects are often described in a parametric form. For performing efficiently on these objects, it can also be interesting to manipulate implicit representations. We consider particular projections techniques based on new resultant constructions or syzygies, which allow to transform parametric representations into implicit ones. These problems can be reformulated in terms of linear algebra. We investigate methods which exploit this matrix representation based on resultant constructions.

They involve structured matrices such as Hankel, Toeplitz, Bezoutian matrices or their generalization in several variables. We investigate algorithms that exploit their properties and their implications in solving polynomial equations.

We are also interested in the "effective" use of duality, that is, the properties of linear forms on the polynomials or quotient rings by ideals. We undertake a detailed study of these tools from an algorithmic perspective, which yields the answer to basic questions in algebraic geometry and brings a substantial improvement on the complexity of resolution of these problems.

We are also interested in subdivision methods, which are able to efficiently localise the real roots of polynomial equations. The specificities of these methods are local behavior, fast convergence properties and robustness. Key problems are related to the analysis of multiple points.

An important issue while developing these methods is to analyse their practical and algorithmic behavior. Our aim is to obtain good complexity bounds and practical efficiency by exploiting the structure of the problem.

3.4. Symbolic numeric analysis

While treating practical problems, noisy data appear and incertitude has to be taken into account. The objective is to devise adapted techniques for analyzing the geometric properties of the algebraic models in this context.

Analysing a geometric model requires tools for structuring it, which first leads to study its singularities and its topology. In many context, the input representation is given with some error so that the analysis should take into account not only one model but a neighborhood of models.

The analysis of singularities of geometric models provides a better understanding of their structures. As a result, it may help us better apprehend and approach modeling problems. We are particularly interested in applying singularity theory to cases of implicit curves and surfaces, silhouettes, shadows curves, moved curves, medial axis, self-intersections, appearing in algorithmic problems in CAGD and shape analysis.

The representation of such shapes is often given with some approximation error. It is not surprising to see that symbolic and numeric computations are closely intertwined in this context. Our aim is to exploit the complementarity of these domains, in order to develop controlled methods.

The numerical problems are often approached locally. However, in many situations it is important to give global answers, making it possible to certify computation. The symbolic-numeric approach combining the algebraic and analytical aspects, intends to address these local-global problems. Especially, we focus on certification of geometric predicates that are essential for the analysis of geometrical structures.

The sequence of geometric constructions, if treated in an exact way, often leads to a rapid complexification of the problems. It is then significant to be able to approximate the geometric objects while controlling the quality of approximation. We investigate subdivision techniques based on the algebraic formulation of our problems which allow us to control the approximation, while locating interesting features such as singularities.

According to an engineer in CAGD, the problems of singularities obey the following rule: less than 20% of the treated cases are singular, but more than 80% of time is necessary to develop a code allowing to treat them correctly. Degenerated cases are thus critical from both theoretical and practical perspectives. To resolve these difficulties, in addition to the qualitative studies and classifications, we also study methods of *perturbations* of symbolic systems, or adaptive methods based on exact arithmetics.

The problem of decomposition and factorisation is also important. We are interested in a new type of algorithms that combine the numerical and symbolic aspects, and are simultaneously more effective and reliable. A typical problem in this direction is the problem of approximate factorization, which requires to analyze perturbations of the data, which enables us to break up the problem.

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

3. Scientific Foundations

3.1. Mesh Generation and Geometry Processing

Meshes are becoming commonplace in a number of applications ranging from engineering to multimedia through biomedecine and geology. For rendering, the quality of a mesh refers to its approximation properties. For numerical simulation, a mesh is not only required to faithfully approximate the domain of simulation, but also to satisfy size as well as shape constraints. The elaboration of algorithms for automatic mesh generation is a notoriously difficult task as it involves numerous geometric components: Complex data structures and algorithms, surface approximation, robustness as well as scalability issues. The recent trend to reconstruct domain boundaries from measurements adds even further hurdles. Armed with our experience on triangulations and algorithms, and with components from the CGAL library, we aim at devising robust algorithms for 2D, surface, 3D mesh generation as well as anisotropic meshes. Our research in mesh generation primarily focuses on the generation of simplicial meshes, i.e. triangular and tetrahedral meshes. We investigate both greedy approaches based upon Delaunay refinement and filtering, and variational approaches based upon energy functionals and associated minimizers.

The search for new methods and tools to process digital geometry is motivated by the fact that previous attempts to adapt common signal processing methods have led to limited success: Shapes are not just another signal but a new challenge to face due to distinctive properties of complex shapes such as topology, metric, lack of global parameterization, non-uniform sampling and irregular discretization. Our research in geometry processing ranges from surface reconstruction to surface remeshing through curvature estimation, principal component analysis, surface approximationand surface mesh parameterization. Another focus is on the robustness of the algorithms to defect-laden data. This focus stems from the fact that acquired geometric data obtained through measurements or designs are rarely usable directly by downstream applications. This generates bottlenecks, i.e., parts of the processing pipeline which are too labor-intensive or too brittle for practitioners. Beyond reliability and theoretical foundations, our goal is to design methods which are also robust to raw, unprocessed inputs.

3.2. Topological and Geometric Inference

Due to the fast evolution of data acquisition devices and computational power, scientists in many areas are asking for efficient algorithmic tools for analyzing, manipulating and visualizing more and more complex shapes or complex systems from approximative data. Many of the existing algorithmic solutions which come with little theoretical guarantee provide unsatisfactory and/or unpredictable results. Since these algorithms take as input discrete geometric data, it is mandatory to develop concepts that are rich enough to robustly and correctly approximate continuous shapes and their geometric properties by discrete models. Ensuring the correctness of geometric estimations and approximations on discrete data is a sensitive problem in many applications.

Data sets being often represented as point sets in high dimensional spaces, there is a considerable interest in analyzing and processing data in such spaces. Although these point sets usually live in high dimensional spaces, one often expects them to be located around unknown, possibly non linear, low dimensional shapes. These shapes are usually assumed to be smooth submanifolds or more generally compact subsets of the ambient space. It is then desirable to infer topological (dimension, Betti numbers,...) and geometric characteristics (singularities, volume, curvature,...) of these shapes from the data. The hope is that this information will help to better understand the underlying complex systems from which the data are generated. In spite of recent promising results, many problems still remain open and to be addressed, need a tight collaboration between mathematicians and computer scientists. In this context, our goal is to contribute to the development of new mathematically well founded and algorithmically efficient geometric tools for data analysis and processing of complex geometric objects. Our main targeted areas of application include machine learning, data mining, statistical analysis, and sensor networks.

3.3. Data Structures and Robust Geometric Computation

GEOMETRICA has a large expertise of algorithms and data structures for geometric problems.We are pursuing efforts to design efficient algorithms from a theoretical point of view, but we also put efforts in the effective implementation of these results.

In the past years, we made significant contributions to algorithms for computing Delaunay triangulations (which are used by meshes in the above paragraph). We are still working on the practical efficiency of existing algorithms to compute or to exploit classical Euclidean triangulations in 2 and 3 dimensions, but the current focus of our research is more aimed towards extending the triangulation efforts in several new directions of research.

One of these directions is the triangulation of non Euclidean spaces such as periodic or projective spaces, with various potential applications ranging from astronomy to granular material simulation.

Another direction is the triangulation of moving points, with potential applications to fluid dynamics where the points represent some particles of some evolving physical material, and to variational methods devised to optimize point placement for meshing a domain with a high quality elements.

Increasing the dimension of space is also a stimulating direction of research, as triangulating points in medium dimension (say 4 to 15) has potential applications and makes new challenges to trade exponential complexity of the problem in the dimension for the possibility to reach effective and practical results in reasonably small dimensions.

On the complexity analysis side, we pursue efforts to obtain complexity analysis in some practical situations involving randomized or stochastic hypotheses. On the algorithm design side, we are looking for new paradigms to exploit parallelism on modern multicore hardware architectures.

Finally, all this work is done while keeping in mind concerns related to effective implementation of our work, practical efficiency and robustness issues which have become a background task of all different works made by GEOMETRICA.

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

3. Scientific Foundations

3.1. Type theory and formalization of mathematics

The calculus of inductive constructions is a branch of type theory that serves as a foundation for theorem proving tools, especially the Coq proof assistant. It is powerful enough to formalize complex mathematics, based on algebraic structures and operations. This is especially important as we want to produce proofs of logical properties for these algebraic structures, a goal that is only marginally addressed in most scientific computation systems.

The calculus of inductive constructions also makes it possible to write algorithms as recursive functional programs which manipulate tree-like data structures. A third important characteristic of this calculus is that it is also a language for manipulating proofs. All this makes this calculus a tool of choice for our investigations. However, this language is still being improved and part of our work concerns these improvements.

3.2. Verification of scientific algorithms

To produce certified algorithms, we use the following approach: instead of attempting to prove properties of an existing program written in a conventional programming language such as C or Java, we produce new programs in the calculus of constructions whose correctness is an immediate consequence of their construction. This has several advantages. First, we work at a high level of abstraction, independently of the target implementation language. Second, we concentrate on specific characteristics of the algorithm, and abstract away from the rest (for instance, we abstract away from memory management or data implementation strategies). Thus, we are able to address more high-level mathematics and to express more general properties without being overwhelmed by implementation details.

However, this approach also presents a few drawbacks. For instance, the calculus of constructions usually imposes that recursive programs should explicitly terminate for all inputs. For some algorithms, we need to use advanced concepts (for instance, well-founded relations) to make the property of termination explicit, and proofs of correctness become especially difficult in this setting.

3.3. Programming language semantics

To bridge the gap between our high-level descriptions of algorithms and conventional programming languages, we investigate the algorithms that are present in programming language implementations, for instance algorithms that are used in a compiler or a static analysis tool. For these algorithms, we generally base our work on the semantic description of a language. The properties that we attempt to prove for an algorithm are, for example, that an optimization respects the meaning of programs or that the programs produced are free of some unwanted behavior. In practice, we rely on this study of programming language semantics to propose extensions to theorem proving tools or to participate in the verification that compilers for conventional programming languages are exempt from bugs.

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

3. Scientific Foundations

3.1. Introduction

Within the extensive field of inverse problems, much of the research by APICS deals with reconstructing solutions of classical elliptic PDEs from their boundary behaviour. Perhaps the most basic example of such a problem is harmonic identification of a stable linear dynamical system: the transfer-function f is holomorphic in the right half-pane, which means it satisfies there the Cauchy-Riemann equation $\overline{\partial} f = 0$, and in principle f can be recovered from its values on the imaginary axis, *e.g.* by Cauchy formula.

Practice is not nearly as simple, for f is only measured pointwise in the pass-band of the system which makes the problem ill-posed [69]. Moreover, the transfer function is usually sought in specific form, displaying the necessary physical parameters for control and design. For instance if f is rational of degree n, it satisfies $\overline{\partial} f = \sum_1^n a_j \delta_{z_j}$ where the z_j are its poles, and finding the domain of holomorphy (*i.e.* locating the z_j) amounts to solve a (degenerate) free-boundary inverse problem, this time on the left half-plane. To address these questions, the team has developed a two-step approach as follows.

Step 1: To determine a complete model, that is, one which is defined for every frequency, in a sufficiently flexible function class (*e.g.* Hardy spaces). This ill-posed issue requires regularization, for instance constraints on the behaviour at non-measured frequencies.

Step 2: To compute a reduced order model. This typically consists of rational approximation of the complete model obtained in step 1, or phase-shift thereof to account for delays. Derivation of the complete model is important to achieve stability of the reduced one.

Step 1 makes connection with extremal problems and analytic operator theory, see section 3.3.1 . Step 2 involves optimization, and some Schur analysis to parametrize transfer matrices of given Mc-Millan degree when dealing with systems having several inputs and output, see section 3.3.2.2 . It also makes contact with the topology of rational functions, to count critical points and to derive bounds, see section 3.3.2 . Moreover, this step raises issues in approximation theory regarding the rate of convergence and whether the singularities of the approximant (*i.e.* its poles) converge to the singularities of the approximated function; this is where logarithmic potential theory becomes effective, see section 3.3.3 .

Iterating the previous steps coupled with a sensitivity analysis yields a tuning procedure which was first demonstrated in [77] on resonant microwave filters.

Similar steps can be taken to approach design problems in frequency domain, replacing measured behaviour by desired behaviour. However, describing achievable responses from the design parameters at hand is generally cumbersome, and most constructive techniques rely on rather specific criteria adapted to the physics of the problem. This is especailly true of circuits and filters, whose design classically appeals to standard polynomial extremal problems and realization procedures from system theory [70], [55]. APICS is active in this field, where we introduced the use of Zolotarev-like problems for microwave multiband filter design. We currently favor interpolation techniques because of their transparency with respect to parameter use, see section 3.2.2 .

In another connection, the example of harmonic identification quickly suggests a generalization of itself. Indeed, on identifying $\mathbb{C}$ with $\mathbb{R}^2$, holomorphic functions become conjugate-gradients of harmonic functions so that harmonic identification is, after all, a special case of a classical issue: to recover a harmonic function on a domain from partial knowledge of the Dirichlet-Neumann data; portion of the boundary where data are not available may be unknown, in which case we meet a free boundary problem. This framework for 2-D non-destructive control was first advocated in [59] and subsequently received considerable attention. This framework makes it clear how to state similar problems in higher dimensions and for more general operators than the Laplacian, provided solutions are essentially determined by the trace of their gradient on part of the boundary which is the case for elliptic equations [1] [79]. All these questions are particular instances of the

so-called inverse potential problem, where a measure μ has to be recovered from knowledge of the gradient of its potential (*i.e.*, the field) on part of a hypersurface (a curve in 2-D) encompassing the support of μ. For Laplace's operator, potentials are logarithmic in 2-D and Newtonian in higher dimensions. For elliptic operators with non constant coefficients, the potential depends on the form of fundamental solutions and is less manageable because it is no longer of convolution type. In any case, by construction, the operator applied to the potential yields back the measure.

Inverse potential problems are severely indeterminate because infinitely many measures within an open set produce the same field outside this set [68]. In step 1 above we implicitly removed this indeterminacy by requiring that the measure be supported on the boundary (because we seek a function holomorphic throughout the right half space), and in step 2 by requiring, say, in case of rational approximation that the measure be discrete in the left half-plane. The same discreteness assumption prevails in 3-D inverse source problems. To recap, the gist of our approach is to approximate boundary data by (boundary traces of) fields arising from potentials of measures with specific support. Note this is different from standard approaches to inverse problems, where descent algorithms are applied to integration schemes of the direct problem; in such methods, it is the equation which gets approximated (in fact: discretized).

Along these lines, the team initiated the use of steps 1 and 2 above, along with singularity analysis, to approach issues of nondestructive control in 2 and 3-D [41] [6], [2]. We are currently engaged in two kinds of generalization, further described in section 3.2.1 . The first one deals with non-constant conductivities, where Cauchy-Riemann equations for holomorphic functions are replaced by conjugate Beltrami equations for pseudo-holomorphic functions; there we seek applications to plasma confinement. The other one lies with inverse source problems for Laplace's equation in 3-D, where holomorphic functions are replaced by harmonic gradients, developing applications to EEG/MEG and inverse magnetization problems in paleomagnetism, see section 4.2

The main approximation-theoretic tools developed by APICS to get to grips with issues mentioned so far are outlined in section 3.3 . In section 3.2 to come, we make more precise which problems are considered and for which applications.

3.2. Range of inverse problems

3.2.1. *Elliptic partial differential equations (PDE)*

Participants: Laurent Baratchart, Slah Chaabi, Juliette Leblond, Ana-Maria Nicu, Dmitry Ponomarev, Elodie Pozzi.

This work is done in collaboration with Alexander Borichev (Univ. Provence).

Reconstructing Dirichlet-Neumann boundary conditions for a function harmonic in a plane domain when these are known on a strict subset E of the boundary, is equivalent to recover a holomorphic function in the domain from its boundary values on E. This is the problem raised on the half-plane in step 1 of section 3.1 . It makes good sense in holomorphic Hardy spaces where functions are determined by their values on boundary subsets of positive linear measure, which is the framework for problem (P) in section 3.3.1 . Such problems naturally arise in nondestructive testing of 2-D (or cylindrical) materials from partial electrical measurements on the boundary. Indeed, the ratio between tangential and normal currents (so-called Robin coefficient) tells about corrosion of the material. Solving problem (P) where ψ is chosen to be the response of some uncorroded piece with identical shape allows one to approach such questions, and this was an initial application of holomorphic extremal problems to non-destructive control [56], [52].

---

[1] There is a subtle difference here between dimension 2 and higher. Indeed, a function holomorphic on a plane domain is defined by its non-tangential limit on a boundary subset of positive linear measure, but there are non-constant harmonic functions in the 3-D ball, C^1 up to the boundary sphere, yet having vanishing gradient on a subset of positive measure of the sphere

A recent application by the team deals with non-constant conductivity over a doubly connected domain, E being the outer boundary. Measuring Dirichlet-Neumann data on E, we wanted to check whether the solution is constant on the inner boundary. We first had to define and study Hardy spaces of the conjugate Beltrami equation, of which the conductivity equation is the compatibility condition (just like Laplace's equation is the compatibility condition of the Cauchy-Riemann system). This was done in references [5] and [35]. Then, solving an obvious modification of problem (P) allows one to numerically check what we want. Further, the value of this extremal problem defines a criterion on inner boundaries, and subsequently a descent algorithm was set up to improve the initial boundary into one where the solution is closer to being constant, thereby trying to solve a free boundary problem..

When the domain is regarded as separating the edge of a tokamak's vessel from the plasma (rotational symmetry makes this a 2-D problem), the procedure just described suits plasma control from magnetic confinement. It was successfully applied in collaboration with CEA (the French nuclear agency) and the University of Nice (JAD Lab.) to data from *Tore Supra* [58], see section 6.2 . This procedure is fast because no numerical integration of the underlying PDE is needed, as an explicit basis of solutions to the conjugate Beltrami equation was found in this case.

Three-dimensional versions of step 1 in section 3.1 are also considered, namely to recover a harmonic function (up to a constant) in a ball or a half-space from partial knowledge of its gradient on the boundary. Such questions arise naturally in connection with neurosciences and medical imaging (electroencephalography, EEG) or in paleomagnetism (analysis of rocks magnetization) [2] [37], see section 6.1 . They are not yet as developed as the 2-D case where the power of complex analysis is at work, but considerable progress was made over the last years through methods of harmonic analysis and operator theory.

The team is also concerned with non-destructive control problems of localizing defaults such as cracks, sources or occlusions in a planar or 3-dimensional domain, from boundary data (which may correspond to thermal, electrical, or magnetic measurements). These defaults can be expressed as a lack of analyticity of the solution of the associated Dirichlet-Neumann problem and we approach them using techniques of best rational or meromorphic approximation on the boundary of the object [4] [16], see sections 3.3.2 and 4.2 . In fact, the way singularities of the approximant relate to the singularities of the approximated function is an all-pervasive theme in approximation theory, and for appropriate classes of functions the location of the poles of a best rational approximant can be used as an estimator of the singularities of the approximated function (see section 6.1). This circle of ideas is much in the spirit of step 2 in section 3.1 .

A genuine 3-dimensional theory of approximation by discrete potentials, though, is still in its infancy.

3.2.2. *Systems, transfer and scattering*

Participants: Laurent Baratchart, Sylvain Chevillard, Sanda Lefteriu, Martine Olivi, Fabien Seyfert.

Through initial contacts with CNES, the French space agency, the team came to work on identification-for-tuning of microwave electromagnetic filters used in space telecommunications (see section 4.3). The problem was to recover, from band-limited frequency measurements, the physical parameters of the device under examination. The latter consists of interconnected dual-mode resonant cavities with negligible loss, hence its scattering matrix is modelled by a 2×2 unitary-valued matrix function on the frequency line, say the imaginary axis to fix ideas. In the bandwidth around the resonant frequency, a modal approximation of the Helmholtz equation in the cavities shows that this matrix is approximately rational, of Mc-Millan degree twice the number of cavities.

This is where system theory enters the scene, through the so-called *realization* process mapping a rational transfer function in the frequency domain to a state-space representation of the underlying system as a system of linear differential equations in the time domain. Specifically, realizing the scattering matrix allows one to construct a virtual electrical network, equivalent to the filter, the parameters of which mediate in between the frequency response and the geometric characteristics of the cavities (*i.e.* the tuning parameters).

Hardy spaces, and in particular the Hilbert space H^2, provide a framework to transform this classical ill-posed issue into a series of well-posed analytic and meromorphic approximation problems. The procedure sketched in section 3.1 now goes as follows:

1. infer from the pointwise boundary data in the bandwidth a stable transfer function (*i.e.* one which is holomorphic in the right half-plane), that may be infinite dimensional (numerically: of high degree). This is done by solving in the Hardy space H^2 of the right half-plane a problem analogous to (P) in section 3.3.1, taking into account prior knowledge on the decay of the response outside the bandwidth, see [18] for details.

2. From this stable model, a rational stable approximation of appropriate degree is computed. For this a descent method is used on the relatively compact manifold of inner matrices of given size and degree, using a novel parametrization of stable transfer functions [18].

3. From this rational model, realizations meeting certain constraints imposed by the technology in use are computed (see section 6.3). These constraints typically come from the nature and topology of the equivalent electrical network used to model the filter. This network is composed of resonators, coupled to each other by some specific coupling topology. Performing this realization step for given coupling topology can be recast, under appropriate compatibility conditions [8], as the problem of solving a zero-dimensional multivariate polynomial system. To tackle this problem in practice, we use Groebner basis techniques as well as continuation methods as implemented in the Dedale-HF software (5.4).

Let us also mention that extensions of classical coupling matrix theory to frequency-dependent (reactive) couplings have lately been carried-out [1] for wide-band design applications, but further study is needed to make them effective.

Subsequently APICS started investigating issues pertaining to filter design rather than identification. Given the topology of the filter, a basic problem is to find the optimal response with respect to amplitude specifications in frequency domain bearing on rejection, transmission and group delay of scattering parameters. Generalizing the approach based on Tchebychev polynomials for single band filters, we recast the problem of multi-band response synthesis in terms of a generalization of classical Zolotarev min-max problem [30] to rational functions [11]. Thanks to quasi-convexity, the latter can be solved efficiently using iterative methods relying on linear programming. These are implemented in the software easy-FF (see section 5.5).

Later, investigations by the team extended to design and identification of more complex microwave devices, like multiplexers and routers, which connect several filters through wave guides. Schur analysis plays an important role in such studies, which is no surprise since scattering matrices of passive systems are of Schur type (*i.e.* contractive in the stability region). The theory originates with the work of I. Schur [76], who devised a recursive test to check for contractivity of a holomorphic function in the disk. Generalizations thereof turned out to be very efficient to parametrize solutions to contractive interpolation problems subject to a well-known compatibility condition (positive definiteness of the so-called Pick matrix) [32]. Schur analysis became quite popular in electrical engineering, as the Schur recursion precisely describes how to chain two-port circuits.

Dwelling on this, members of the team contributed to differential parametrizations (atlases of charts) of loss-less matrix functions to the theory [31][12], [10]. They are of fundamental use in our rational approximation software RARL2 (see section 5.1). Schur analysis is also instrumental to approach de-embedding issues considered in section 6.4, and provides further background to current studies by the team of synthesis and adaptation problems for multiplexers. At the heart of the latter lies a variant of contractive interpolation with degree constraint introduced in [62].

We also mention the role played by multipoint Schur analysis in the team's investigation of spectral representation for certain non-stationary discrete stochastic processes [3], [36].

Recently, in collaboration with UPV (Bilbao), our attention was driven by CNES, to questions of stability relative to high-frequency amplifiers, see section 7.2. Contrary to previously mentioned devices, these are *active* components. The amplifier can be linearized at a functioning point and admittances of the corresponding electrical network can be computed at various frequencies, using the so-called harmonic balance method. The

goal is to check for stability of this linearised model. The latter is composed of lumped electrical elements namely inductors, capacitors, negative *and* positive reactors, transmission lines, and commanded current sources. Research so far focused on determining the algebraic structure of admittance functions, and setting up a function-theoretic framework to analyse them. In particular, much effort was put on realistic assumptions under which a stable/unstable decomposition can be claimed in $H^2 \oplus \overline{H^2}$ (see section 6.5). Under them, the unstable part of the elements under examination is rational and we expect to bring valuable estimates of stability to the designer using the general scheme in section 3.1 .

3.3. Approximation of boundary data

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Martine Olivi, Dmitry Ponomarev, Elodie Pozzi, Fabien Seyfert.

The following people are collaborating with us on these topics: Bernard Hanzon (Univ. Cork, Ireland), Jean-Paul Marmorat (Centre de mathématiques appliquées (CMA), École des Mines de Paris), Jonathan Partington (Univ. Leeds, UK), Ralf Peeters (Univ. Maastricht, NL), Edward Saff (Vanderbilt University, Nashville, USA), Herbert Stahl (TFH Berlin), Maxim Yattselev (Univ. Oregon at Eugene, USA).

3.3.1. *Best constrained analytic approximation*

In dimension 2, the prototypical problem to be solved in step 1 of section 3.1 may be described as: given a domain $D \subset \mathbb{R}^2$, we want to recover a holomorphic function from its values on a subset of the boundary of D. Using conformal mapping, it is convenient for the discussion to normalize D. So, in the simply connected case, we fix D to be the unit disk with boundary the unit circle T. We denote by H^p the Hardy space of exponent p which is the closure of polynomials in the L^p-norm on the circle if $1 \le p < \infty$ and the space of bounded holomorphic functions in D if $p = \infty$. Functions in H^p have well-defined boundary values in $L^p(T)$, which makes it possible to speak of (traces of) analytic functions on the boundary.

To find an analytic function in D approximately matching measured values f on a sub-arc K of T, we formulate a constrained best approximation problem as follows.

> (P) Let $1 \le p \le \infty$, K a sub-arc of T, $f \in L^p(K)$, $\psi \in L^p(T \smallsetminus K)$ and $M > 0$; find a function $g \in H^p$ such that $\|g - \psi\|_{L^p(T \smallsetminus K)} \le M$ and $g - f$ is of minimal norm in $L^p(K)$ under this constraint.

Here ψ is a reference behaviour capturing *a priori* assumptions on the behaviour of the model off K, while M is some admissible deviation from them. The value of p reflects the type of stability which is sought and how much one wants to smoothen the data. The choice of L^p classes is well-adapted to handling pointwise measurements.

To fix terminology we refer to (P) as a *bounded extremal problem*. As shown in [40], [42], [47], for $1 < p \le \infty$, the solution to this convex infinite-dimensional optimization problem can be obtained upon iterating with respect to a Lagrange parameter the solution to spectral equations for some appropriate Hankel and Toeplitz operators. These equations in turn involve the solution to the standard extremal problem below best approximation problem (P_0) below [61]:

> (P_0) Let $1 \le p \le \infty$ and $\varphi \in L^p(T)$; find a function $g \in H^p$ such that $g - \varphi$ is of minimal norm in $L^p(T)$.

The case $p = 1$ of (P) is essentially open.

Various modifications of (P) have been studied in order to meet specific needs. For instance when dealing with loss-less transfer functions (see section 4.3), one may want to express the constraint on $T \smallsetminus K$ in a pointwise manner: $|g - \psi| \le M$ a.e. on $T \smallsetminus K$, see [43]. In this form, it comes close to (but still is different from) H^∞ frequency optimization methods for control [64], [75].

The analog of problem (P) on an annulus, K being now the outer boundary, can be seen as a means to regularize a classical inverse problem occurring in nondestructive control, namely recovering a harmonic function on the inner boundary from Dirichlet-Neumann data on the outer boundary (see sections 3.2.1 , 4.2 , 6.1.1 , 6.2). For $p = 2$ the solution is analysed in [66]. It may serve as a tool to approach Bernoulli type problems where we are given data on the outer boundary and we *seek the inner boundary*, knowing it is a level curve of the flux. Then, the Lagrange parameter indicates which deformation should be applied on the inner contour in order to improve data fitting.

This is discussed in sections 3.2.1 and 6.2 for more general equations than the Laplacian, namely isotropic conductivity equations of the form $\text{div}(\sigma \nabla u) = 0$ where σ is non constant. In this case Hardy spaces in problem (P) become those of a so-called conjugate or real Beltrami equation [65], which are studied for $1 < p < \infty$ in [5], [35]. Expansions of solutions needed to constructively handle such issues have been carried out in [58].

Though originally considered in dimension 2, problem (P) carries over naturally to higher dimensions where analytic functions get replaced by gradients of harmonic functions. Namely, given some open set $\Omega \subset \mathbb{R}^n$ and a $\mathbb{R}^n$-valued vector V field on an open subset O of the boundary of Ω, we seek a harmonic function in Ω whose gradient is close to V on O.

When Ω is a ball or a half-space, a convenient substitute of holomorphic Hardy spaces is provided by Stein-Weiss Hardy spaces of harmonic gradients [78]. Conformal maps are no longer available in $\mathbb{R}^n$ for $n > 2$ and other geometries have not been much studied so far. On the ball, the analog of problem (P) is

> $(P1)$ Let $1 \leq p \leq \infty$ and $B \subset \mathbb{R}^n$ the unit ball. Fix O an open subset of the unit sphere $S \subset \mathbb{R}^n$. Let further $V \in L^p(O)$ and $W \in L^p(S \smallsetminus O)$ be $\mathbb{R}^n$-valued vector fields, and $M > 0$; find a harmonic gradient $G \in H^p(B)$ such that $\|G - W\|_{L^p(S \smallsetminus O)} \leq M$ and $G - V$ is of minimal norm in $L^p(O)$ under this constraint.

When $p = 2$, spherical harmonics offer a reasonable substitute to Fourier expansions and problem $(P1)$ was solved in [2], together with its natural analog on a shell. The solution generalizes the Toeplitz operator approach to bounded extremal problems [40], and constructive aspects of the procedure (harmonic 3-D projection, Kelvin and Riesz transformation, spherical harmonics) were derived. An important ingredient is a refinement of the Hodge decomposition allowing us to express a $\mathbb{R}^n$-valued vector field in $L^p(S)$, $1 < p < \infty$, as the sum of a vector field in $H(B)$, a vector field in $H^p(\mathbb{R}^n \smallsetminus \overline{B})$, and a tangential divergence free vector field. If $p = 1$ or $p = \infty$, L^p must be replaced respectively by the real Hardy space H^1 and the bounded mean oscillation space BMO, and H^∞ should be modified accordingly. This decomposition was fully discussed in [37] (for the case of the half-space) where it plays a fundamental role.

Problem $(P1)$ is still under investigation in the case $p = \infty$, where even the case where $O = S$ is pending because a substitute of the Adamjan-Arov-Krein theory [72] is still to be built in dimension greater than 2.

Such problems arise in connection with source recovery in electro/mgneto encephalography and paleomagnetism, as discussed in sections 3.2.1 and 4.2 .

3.3.2. *Best meromorphic and rational approximation*

The techniques explained in this section are used to solve step 2 in section 3.2*via* conformal mapping and subsequently instrumental to approach inverse boundary value problems for Poisson equation $\Delta u = \mu$, where μ is some (unknown) distribution.

3.3.2.1. Scalar meromorphic and rational approximation

Let as before D designate the unit disk, and T the unit circle. We further put R_N for the set of rational functions with at most N poles in D, which allows us to define the meromorphic functions in $L^p(T)$ as the traces of functions in $H^p + R_N$.

A natural generalization of problem (P_0) is:

> (P_N) Let $1 \leq p \leq \infty$, $N \geq 0$ an integer, and $f \in L^p(T)$; find a function $g_N \in H^p + R_N$ such that $g_N - f$ is of minimal norm in $L^p(T)$.

Only for $p = \infty$ and continuous f it is known how to solve (P_N) in closed form. The unique solution is given by AAK theory (named after Adamjan, Arov and Krein), that connects the spectral decomposition of Hankel operators with best approximation in Hankel norm [72]. This theory allows one to express g_N in terms of the singular vectors of the Hankel operator with symbol f. The continuity of g_N as a function of f only holds for stronger norms than uniform.

The case $p = 2$ is of special importance. In particular when $f \in \overline{H}^2$, the Hardy space of exponent 2 of the *complement* of D in the complex plane (by definition, $h(z)$ belongs to $\overline{H}^p$ if, and only if, $h(1/z)$ belongs to H^p), then (P_N) reduces to rational approximation. Moreover, it turns out that the associated solution $g_N \in R_N$ has no pole outside D, hence it is a *stable* rational approximant to f. However, in contrast with the situation when $p = \infty$, this approximant may *not* be unique.

The former Miaou project (predecessor of Apics) has designed an adapted steepest-descent algorithm for the case $p = 2$ whose convergence to a *local minimum* is guaranteed; until now it seems to be the only procedure meeting this property. Roughly speaking, it is a gradient algorithm that proceeds recursively with respect to the order N of the approximant, in a compact region of the parameter space [34]. Although it has proved effective in all applications carried out so far (see sections 4.2 , 4.3), it is not known whether the absolute minimum can always be obtained by choosing initial conditions corresponding to *critical points* of lower degree (as is done by the RARL2 software, section 5.1).

In order to establish global convergence results, APICS has undertook a long-term study of the number and nature of critical points, in which tools from differential topology and operator theory team up with classical approximation theory. The main discovery is that the nature of the critical points (*e.g.*, local minima, saddles...) depends on the decrease of the interpolation error to f as N increases [44]. Based on this, sufficient conditions have been developed for a local minimum to be unique. These conditions are hard to use in practice because they require strong estimates of the approximation error. These are often difficult to obtain for a given function, and are usually only valid for large N. Examples where uniqueness or asymptotic uniqueness has been proved this way include transfer functions of relaxation systems (*i.e.* Markov functions) [48] and more generally Cauchy integrals over hyperbolic geodesic arcs [50] and certain entire functions [46].

An analog to AAK theory has been carried out for $2 \leq p < \infty$ [47]. Although not computationally as powerful, it can be used to derive lower bounds and helps analysing the behaviour of poles. When $1 \leq p < 2$, problem (P_N) is still fairly open.

A common feature to all these problems is that critical point equations express non-Hermitian orthogonality relations for the denominator of the approximant. This makes connection with interpolation theory [51][7] and is used in an essential manner to assess the behaviour of the poles of the approximants to functions with branchpoint-type singularities, which is of particular interest for inverse source problems (*cf.* sections 5.6 and 6.1).

In higher dimensions, the analog of problem (P_N) is the approximation of a vector field with gradients of potentials generated by N point masses instead of meromorphic functions. The issue is by no means understood at present, and is a major endeavour of future research problems.

Certain constrained rational approximation problems, of special interest in identification and design of passive systems, arise when putting additional requirements on the approximant, for instance that it should be smaller than 1 in modulus. Such questions have become over years an increasingly significant part of the team's activity (see section 4.3). For instance, convergence properties of multipoint Schur approximants, which are rational interpolants preserving contractivity of a function, were analysed in [3]. Such approximants are useful in prediction theory of stochastic processes, but since they interpolate inside the domain of holomorphy they are of limited use in frequency design.

In another connection, the generalization to several arcs of classical Zolotarev problems [74] is an achievement by the team which is useful for multiband synthesis [11]. Still, though the modulus of the response is the first concern in filter design, variation of the phase must nevertheless remain under control to avoid unacceptable distortion of the signal. This specific but important issue has less structure and was approached

using constrained optimization; a dedicated code has been developed under contract with the CNES (see section 5.5).

3.3.2.2. Matrix-valued rational approximation

Matrix-valued approximation is necessary for handling systems with several inputs and outputs, and it generates substantial additional difficulties with respect to scalar approximation, theoretically as well as algorithmically. In the matrix case, the McMillan degree (*i.e.* the degree of a minimal realization in the System-Theoretic sense) generalizes the degree.

The problem we want to consider reads: *Let $\mathcal{F} \in (H^2)^{m \times l}$ and n an integer; find a rational matrix of size $m \times l$ without poles in the unit disk and of McMillan degree at most n which is nearest possible to $\mathcal{F}$ in $(H^2)^{m \times l}$.* Here the L^2 norm of a matrix is the square root of the sum of the squares of the norms of its entries.

The scalar approximation algorithm [34], mentioned in section 3.3.2.1 , generalizes to the matrix-valued situation [60]. The first difficulty here consists in the parametrization of transfer matrices of given McMillan degree n, and the inner matrices (*i.e.* matrix-valued functions that are analytic in the unit disk and unitary on the circle) of degree n. The latter enter the picture in an essential manner as they play the role of the denominator in a fractional representation of transfer matrices (using the so-called Douglas-Shapiro-Shields factorization). The set of inner matrices of given degree has the structure of a smooth manifold that allows one to use differential tools as in the scalar case. In practice, one has to produce an atlas of charts (parametrization valid in a neighborhood of a point), and we must handle changes of charts in the course of the algorithm. Such parametrization can be obtained from interpolation theory and Schur type algorithms, the parameters being interpolation vectors or matrices ([31], [10], [12]). Some of them are particularly interesting to compute realizations and achieve filter synthesis ([10] [12]). Rational approximation software codes have been developed in the team (see sections 5.1).

Difficulties relative to multiple local minima naturally arise in the matrix-valued case as well, and deriving criteria that guarantee uniqueness is even more difficult than in the scalar case. The case of rational functions of sought degree or small perturbations thereof (the consistency problem) was solved in   [45]. The case of matrix-valued Markov functions, the first example beyond rational functions, was treated in [33].

Let us stress that the algorithms mentioned above are first to handle rational approximation in the matrix case in a way that converges to local minima, while meeting stability constraints on the approximant.

3.3.3. *Behavior of poles of meromorphic approximants*
Participant: Laurent Baratchart.

The following people collaborate with us on this subject: Herbert Stahl (TFH Berlin), Maxim Yattselev (Univ. Oregon at Eugene, USA).

We refer here to the behaviour of poles of best meromorphic approximants, in the L^p-sense on a closed curve, to functions f defined as Cauchy integrals of complex measures whose support lies inside the curve. If one normalizes the contour to be the unit circle T, we are back to the framework of section 3.3.2.1 and to problem (P_N); invariance of the problem under conformal mapping was established in [6]. Research so far has focused on functions whose singular set inside the contour is zero or one-dimensional.

Generally speaking, the behaviour of poles is particularly important in meromorphic approximation to obtain error rates as the degree goes large and to tackle constructive issues like uniqueness. As explained in section 3.2.1 , we consider this issue in connection with approximation of the solution to a Dirichlet-Neumann problem, so as to extract information on the singularities. The general theme is thus *how do the singularities of the approximant reflect those of the approximated function?* This approach to inverse problem for the 2-D Laplacian turns out to be attractive when singularities are zero- or one-dimensional (see section 4.2). It can be used as a computationally cheap initialization of more precise but heavier numerical optimizations.

As regards crack detection or source recovery, the approach in question boils down to analysing the behaviour of best meromorphic approximants of a function with branch points. For piecewise analytic cracks, or in the case of sources, We were able to prove ([6], [38], [14]) that the poles of the approximants accumulate on some extremal contour of minimum weighted energy linkings the singular points of the crack, or the sources [41]. Moreover, the asymptotic density of the poles turns out to be the Green equilibrium distribution of this contour in D, hence puts heavy charge around the singular points (in particular at the endpoints) which are therefore well localized if one is able to approximate in sufficiently high degree (this is where the method could fail).

The case of two-dimensional singularities is still an outstanding open problem.

It is interesting that inverse source problems inside a sphere or an ellipsoid in 3-D can be attacked with the above 2-D techniques, as applied to planar sections (see section 6.1).

3.3.4. *Miscellaneous*

Participant: Sylvain Chevillard.

Sylvain Chevillard, joined team in November 2010. His coming resulted in APICS hosting a research activity in certified computing, centered around the software *Sollya* of which S. Chevillard is a co-author, see section 5.7 . On the one hand, Sollya is an Inria software which still requires some tuning to a growing community of users. On the other hand, approximation-theoretic methods at work in Sollya are potentially useful for certified solutions to constrained analytic problems described in section 3.3.1 . However, developing Solya is not a long-term objective of APICS.

CASTOR Team

3. Scientific Foundations

3.1. Plasma Physics

Participants: Hervé Guillard, Boniface Nkonga, Afeintou Sangam, Richard Pasquetti, Audrey Bonnement, Marie Martin, Cédric Lachat, Laure Combe, Jacques Blum, Cédric Boulbe, Sebastian Minjeaud.

In order to fulfil the increasing demand, alternative energy sources have to be developed. Indeed, the current rate of fossil fuel usage and its serious adverse environmental impacts (pollution, greenhouse gas emissions, ...) lead to an energy crisis accompanied by potentially disastrous global climate changes.

Controlled fusion power is one of the most promising alternatives to the use of fossil resources, potentially with a unlimited source of fuel. France with the ITER (http://www.iter.org/default.aspx) and Laser Megajoule (http://www-lmj.cea.fr/) facilities is strongly involved in the development of these two parallel approaches to master fusion that are magnetic and inertial confinement. Although the principles of fusion reaction are well understood from nearly sixty years, (the design of tokamak dates back from studies done in the '50 by Igor Tamm and Andreï Sakharov in the former Soviet Union), the route to an industrial reactor is still long and the application of controlled fusion for energy production is beyond our present knowledge of related physical processes. In magnetic confinement, beside technological constraints involving for instance the design of plasma-facing component, one of the main difficulties in the building of a controlled fusion reactor is the poor confinement time reached so far. This confinement time is actually governed by turbulent transport that therefore determines the performance of fusion plasmas. The prediction of the level of turbulent transport in large machines such as ITER is therefore of paramount importance for the success of the researches on controlled magnetic fusion.

The other route for fusion plasma is inertial confinement. In this latter case, large scale hydrodynamical instabilities prevent a sufficient large energy deposit and lower the return of the target. Therefore, for both magnetic and inertial confinement technologies, the success of the projects is deeply linked to the theoretical understanding of plasma turbulence and flow instabilities as well as to mathematical and numerical improvements enabling the development of predictive simulation tools.

3.2. Turbulence Modelling

Participants: Alain Dervieux, Boniface Nkonga, Richard Pasquetti.

Fluid turbulence has a paradoxical situation in science. The Navier-Stokes equations are an almost perfect model that can be applied to any flow. However, they cannot be solved for any flow of direct practical interest. Turbulent flows involve instability and strong dependence to parameters, chaotic succession of more or less organised phenomena, small and large scales interacting in a complex manner. It is generally necessary to find a compromise between neglecting a huge number of small events and predicting more or less accurately some larger events and trends.

In this direction, CASTOR wishes to contribute to the progress of methods for the prediction of fluid turbulence. Taking benefit of its experience in numerical methods for complex applications, CASTOR works out models for predicting flows around complex obstacles, that can be moved or deformed by the flow, and involving large turbulent structures. Taking into account our ambition to provide also short term methods for industrial problems, we consider methods applying to high Reynolds flows, and in particular, methods hybridizing Large Eddy Simulation (LES) with Reynolds Averaging.

Turbulence is the indirect cause of many other phenomena. Fluid-structure interaction is one of them, and can manifest itself for example in Vortex Induced Motion or Vibration. These phenomena can couple also with liquid-gas interfaces and bring new problems. Of particular interest is also the study of turbulence generated noise. In this field, though acoustic phenomena can also in principle be described by the Navier-Stokes equations, they are not generally numerically solved by flow solvers but rather by specialized linear and nonlinear acoustic solvers. An important question is the investigation of the best way to combine a LES simulation with the acoustic propagation of the waves it produces.

3.3. Astrophysical and Environmental flows

Participants: Hervé Guillard, Boniface Nkonga, Sebastian Minjeaud.

Although it seems inappropriate to address the modeling of experimental devices of the size of a tokamak and for instance, astrophysical systems with the same mathematical and numerical tools, it has long been recognized that the behaviour of these systems have a profound unity. This has for consequence for instance that any large conference on plasma physics includes sessions on astrophysical plasmas as well as sessions on laboratory plasmas. CASTOR does not intend to consider fluid models coming from Astrophysics or Environmental flows for themselves. However, the team is interested in the numerical approximation of some problems in this area as they provide interesting reduced models for more complex phenomena. To be more precise, let us give some concrete examples : The development of Rossby waves [1] a common problem in weather prediction has a counterpart in the development of magnetic shear induced instabilities in tokamaks and the understanding of this latter type of instabilities has been largely improved by the Rossby wave model. A second example is the water bag model of plasma physics that has a lot in common with multi-layer shallow water system.

To give a last example, we can stress that the development of the so-called well-balanced finite volume schemes used nowadays in many domains of mathematical physics or engineering was largely motivated by the desire to suppress some problems appearing in the approximation of the shallow water system.

Our goal is therefore to use astrophysical or geophysical models to investigate some numerical questions in contexts that, in contrast with plasma physics or fluid turbulence, do not require huge three dimensional computations but are still of interest for themselves and not only as toy models.

---

[1]Rossby waves are giant meanders in high altitude wind that have major influence on weather. Oceanic Rossby waves are also know to exist and to affect the world ocean circulation

COFFEE Project-Team

3. Scientific Foundations

3.1. Scientific Foundations

Mathematical modeling and computer simulation are among the main research tools for environmental management, risks evaluation and sustainable development policy. Many aspects of the computer codes as well as the PDEs systems on which these codes are based can be considered as questionable regarding the established standards of applied mathematical modeling and numerical analysis. This is due to the intricate multiscale nature and tremendous complexity of those phenomena that require to set up new and appropriate tools. Our research group aims to contribute to bridging the gap by developing advanced abstract mathematical models as well as related computational techniques.

The scientific basis of the proposal is two–fold. On the one hand, the project is "technically–driven": it has a strong content of mathematical analysis and design of general methodology tools. On the other hand, the project is also "application–driven": we have identified a set of relevant problems motivated by environmental issues, which share, sometimes in a unexpected fashion, many common features. The proposal is precisely based on the conviction that these subjects can mutually cross-fertilize and that they will both be a source of general technical developments, and a relevant way to demonstrate the skills of the methods we wish to design.

To be more specific:

- We consider evolution problems describing highly heterogeneous flows (with different phases or with high density ratio). In turn, we are led to deal with non linear systems of PDEs of convection and/or convection–diffusion type.

- The nature of the coupling between the equations can be two–fold, which leads to different difficulties, both in terms of analysis and conception of numerical methods. For instance, the system can couple several equations of different types (elliptic/parabolic, parabolic/hyperbolic, parabolic or elliptic with algebraic constraints, parabolic with degenerate coefficients....). Furthermore, the unknowns can depend on different sets of variables, a typical example being the fluid/kinetic models for particulate flows. In turn, the simulation cannot use a single numerical approach to treat all the equations. Instead, hybrid methods have to be designed which raise the question of fitting them in an appropriate way, both in terms of consistency of the discretization and in terms of stability of the whole computation. For the problems under consideration, the coupling can also arises through interface conditions. It naturally occurs when the physical conditions are highly different in subdomains of the physical domain in which the flows takes place. Hence interface conditions are intended to describe the exchange (of mass, energy...) between the domains. Again it gives rise to rather unexplored mathematical questions, and for numerics it yields the question of defining a suitable matching at the discrete level, that is requested to preserve the properties of the continuous model.

- By nature the problems we wish to consider involve many different scales (of time or length basically). It raises two families of mathematical questions. In terms of numerical schemes, the multiscale feature induces the presence of stiff terms within the equations, which naturally leads to stability issues. A clear understanding of scale separation helps in designing efficient methods, based on suitable splitting techniques for instance. On the other hand asymptotic arguments can be used to derive hierarchy of models and to identify physical regimes in which a reduced set of equations can be used.

MCTAO Team

3. Scientific Foundations

3.1. Control Systems

Our effort is directed toward efficient methods for the *control* of real (physical) systems, based on a *model* of the system to be controlled. *System* refers to the physical plant or device, whereas *model* refers to a mathematical representation of it.

We mostly investigate nonlinear systems whose nonlinearities admit a strong structure derived from the physics; the equations governing their behavior is then rather well known, and the modeling part consists in choosing what phenomena are to be retained in the model used for control design; the other phenomena being treated as perturbations; a more complete model may be used for simulations, for instance. We focus on systems that admit a reliable finite-dimensional model, in continuous time; this means that models are ordinary differential equations, often nonlinear.

Choosing accurate models yet simple enough to allow control design is in itself a key issue; however, modeling or identification as a theory is not per se in the scope of our project.

The extreme generality and versatility of linear control do not contradict the often heard sentence "most real life systems are nonlinear". Indeed, for many control problems, a linear model is sufficient to capture the important features for control. The reason is that most control objectives are local, first order variations around an operating point or a trajectory are governed by a linear control model, and except in degenerate situations (non-controllability of this linear model), the local behavior of a nonlinear dynamic phenomenon is dictated by the behavior of first order variations. Linear control is the hard core of control theory and practice; it has been pushed to a high degree of achievement –see for instance some classics: [56], [42]– that leads to big successes in industrial applications (PID, Kalman filtering, frequency domain design, H^∞ robust control, etc...). It must be taught to future engineers, and it is still a topic of ongoing research.

Linear control by itself however reaches its limits in some important situations:

1. **Non local control objectives.** For instance, steering the system from a region to a reasonably remote other one (path planning and optimal control); in this case, local linear approximation cannot be sufficient.
   It is also the case when some domain of validity (e.g. stability) is prescribed and larger than the region where the linear approximation is dominant.

2. **Local control at degenerate equilibria.** Linear control yields local stabilization of an equilibrium point based on the tangent linear approximation if the latter is controllable. When it is *not*, and this occurs in some physical systems at interesting operating points, linear control is irrelevant and specific nonlinear techniques have to be designed.
   This is in a sense an extreme case of the second paragraph in point 1 : the region where the linear approximation is dominant vanishes.

3. **Small controls.** In some situations, actuators only allow a very small magnitude of the effect of control compared to the effect of other phenomena. Then the behavior of the system without control plays a major role and we are again outside the scope of linear control methods.

4. **Local control around a trajectory.** Sometimes a trajectory has been selected (this appeals to point 1), and local regulation around this reference is to be performed. Linearization in general yields, when the trajectory is not a single equilibrium point, a *time-varying* linear system. Even if it is controllable, time-varying linear systems are not in the scope of most classical linear control methods, and it is better to incorporate this local regulation in the nonlinear design, all the more so as the linear approximation along optimal trajectories is, by nature, often non controllable.

Let us discuss in more details some specific problems that we are studying or plan to study: classification and structure of control systems in section 3.2 , optimal control, and its links with feedback, in section 3.3 , the problem of optimal transport in section 3.4 , and finally problems relevent to a specific class of systems where the control is "small" in section 3.5 .

3.2. Structure of nonlinear control systems

In most problems, choosing the adapted coordinates, or the right quantities that describe a phenomenon, sheds light on a path to the solution. In control systems, it is often crucial to analyze the structure of the model, deduced from physical principles, of the plant of be controlled; this may lead to putting it via some transformations in a simpler form, or a form that is most suitable for control design. For instance, equivalence to a linear system may allow to use linear control; also, the so-called "flatness" property drastically simplifies path planning [48], [63].

A better understanding of the "set of nonlinear models", partly classifying them, has another motivation than facilitating control design for a given system and its model: it may also be a necessary step towards a theory of "nonlinear identification" and modeling. Linear identification is a mature area of control science; its success is mostly due to a very fine knowledge of the structure of the class of linear models: similarly, any progress in the understanding of the structure of the class of nonlinear models would be a contribution to a possible theory of nonlinear identification.

These topics are central in control theory, but raise very difficult mathematical questions: static feedback classification is a geometric problem feasible in principle, although describing invariants explicitly is technically very difficult; and conditions for dynamic feedback equivalence and linearization raise unsolved mathematical problems, that make one wonder about decidability [1].

3.3. Optimal control and feedback control, stabilization

3.3.1. *Optimal control.*

Mathematically speaking, optimal control is the modern branch of the calculus of variations, rather well established and mature [24], [60], [34], [72]. Relying on Hamiltonian dynamics is now prevalent, instead of the standard Lagrangian formalism of the calculus of variations. Also, coming from control engineering, constraints on the control (for instance the control is a force or a torque, naturally bounded) or the state (for example in the shuttle atmospheric re-entry problem there is a constraint on the thermal flux) are imposed; the ones on the state are usual but these on the state yield more complicated necessary optimality conditions and an increased intrinsic complexity of the optimal solutions. Also, in the modern treatment, adapted numerical schemes have to be derived for effective computations of the optimal solutions.

What makes optimal control an applied field is the necessity of computing these optimal trajectories, or rather the controls that produce these trajectories (or, of course, close-by trajectories). Computing a given optimal trajectory and its control as a function of time is a demanding task, with non trivial numerical difficulties: roughly speaking, the Pontryagin Maximum Principle gives candidate optimal trajectories as solutions of a two point boundary value problem (for an ODE) which can be analyzed using mathematical tools from geometric control theory or solved numerically using shooting methods. Obtaining the *optimal synthesis* –the optimal control as a function of the state– is of course a more intricate problem [34], [37]. On the other hand the *value function* –the minimum of the criteria at each point– would give the solution to both questions if it were differentiable, but it is usually not, and is only a viscosity solution of the Hamilton-Jacobi-Bellman (**HJB**) equation (a PDE), whose study requires sophisticated non-smooth and singularity analysis.

---

[1]Consider the simple system with state $(x, y, z) \in I\!R^3$ and two controls that reads $\dot{z} = (\dot{y} - z\dot{x})^2\dot{x}$ after elimination of the controls; it is not known whether it is equivalent to a linear system, or flat; this is because the property amounts to existence of a formula giving the general solution as a function of two arbitrary functions of time and their derivatives up to a certain order, but no bound on this order is known a priori, even for this very particular example.

These questions are not only academic for minimizing a cost is *very* relevant in many control engineering problems. However, modern engineering textbooks in nonlinear control systems like the "best-seller" [51] hardly mention optimal control, and rather put the emphasis on designing a feedback control, as regular and explicit as possible, satisfying some qualitative (and extremely important!) objectives: disturbance attenuation, decoupling, output regulation or stabilization. Optimal control is sometimes viewed as disconnected from automatic control... we shall come back to this unfortunate point.

3.3.2. Feedback, control Lyapunov functions, stabilization.

A control Lyapunov function **(CLF)** is a function that can be made a Lyapunov function (roughly speaking, a function that decreases along all trajectories, some call this an "artificial potential") for the closed-loop system corresponding to *some* feedback law. This can be translated into a partial differential relation sometimes called "Artstein's (in)equation" [27]. There is a definite parallel between a CLF for stabilization, solution of this differential inequation on the one hand, and the value function of an optimal control problem for the system, solution of a HJB equation on the other hand. Now, optimal control is a quantitative objective while stabilization is a qualitative objective; it is not surprising that Artstein (in)equation is very under-determined and has many more solutions than HJB equation, and that it may (although not always) even have smooth ones.

We have, in the team, a longstanding research record on the topic, construction of CLFs and stabilizing feedback controls. This is all the more interesting as our line of research has been pointing in almost opposite directions. [43], [52], [53], [67], [68], [70], [71], [44] insist on the construction of continuous feedback, hence smooth CLFs whereas, on the contrary, [41], [74], [75], [76], [77] proceed with a very fine study of non-smooth CLFs, yet good enough (semi-concave) that they can produce a reasonable discontinuous feedback with reasonable properties.

3.4. Optimal Transport

The study of optimal mass transport problems in the Euclidean or Riemannian setting has a long history which goes from the pioneer works of Monge [65] and Kantorovitch [57] to the recent revival initiated by fundamental contributions due to Brenier [38], [39] and McCann [64]. However, the study of the same transportation problems in the presence of non-holonomic constraints –(like being an admissible trajectory for a control system– is quite new. The first contributors were Ambrosio and Rigot [25] who proved the existence and uniqueness of an optimal transport map for the Monge problem associated with the squared canonical sub-Riemannian distance on the Heisenberg groups. This result was extended later by Agrachev and Lee [22], then by Figalli and Rifford [46] who showed that the Ambrosio-Rigot theorem holds indeed true on many sub-Riemannian manifolds satisfying reasonable assumptions. The problem of existence and uniqueness of an optimal transport map for the squared sub-Riemannian distance on a general complete sub-Riemannian manifold remains open; it is strictly related to the regularity of the sub-Riemannian distance in the product space, and remains a formidable challenge. Generalized notions of Ricci curvatures (bounded from below) in metric spaces have been developed recently by Lott and Villani [61] and Sturm [79], [80]. A pioneer work has been work in the Heisenberg group by Juillet [54] who captured the right notion of curvature in this setting. Agrachev and Lee [23] have elaborated on this work to define new notions of curvatures in three dimensional sub-Riemannian structures. The optimal transport approach happened to be very fruitful in this context. Many things remain to do in a more general context.

One of the results of A. Hindawi's PhD under the supervision of L. Rifford and J.-B. Pomet was to extend regularity theory established in the Euclidean case to the more general quadratic costs associated with linear optimal control problems (LQR), see [50]. This successful result opens a new range of optimal transport problems associated with cost coming from optimal control problems. We can nowadays expect regularity properties for optimal transport maps associated with reasonable optimal control problems with constraints on the state or on the velocities.

We believe that matching optimal transport with geometric control theory is one originality of our team. We expect interactions in both ways.

3.5. Small controls and conservative systems, averaging

Using averaging techniques to study small perturbations of integrable Hamiltonian systems dates back to H. Poincaré or earlier; it gives an approximation of the (slow) evolution of quantities that are preserved in the non-perturbed system. It is very subtle in the case of multiple periods but more elementary in the single period case, here it boils down to taking the average of the perturbation along each periodic orbit; see for instance [26], [78].

When the "perturbation" is a control, these techniques may be used after deciding how the control will depend on time and state and other quantities, for instance it may be used after applying the Pontryagin Maximum Principle as in [30], [31], [40], [49]. Without deciding the control a priori, an "average control system" may be defined as in [2].

The focus is then on studying into details this simpler "averaged" problem, that can often be described by a Riemannian metric for quadratic costs or by a Finsler metric for costs lime minimum time.

This line of research stemmed out of applications to space engineering, see section 4.1 . For orbit transfer in the two-body problem, an important contribution was made by B. Bonnard, J.-B. Caillau and J. Gergaud [31] in explicitly computing the solutions of the average system obtained after applying Pontryagin Maximum Principle to minimizing a quadratic integral cost; this yields an explicit calculation of the optimal control law itself. Studying the Finsler metric issued form the time-minimal case is in progress.

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

3. Scientific Foundations

3.1. High order discretization methods

The applications in computational electromagnetics and computational geoseismics that are considered by the team lead to the numerical simulation of wave propagation in heterogeneous media or/and involve irregularly shaped objects or domains. The underlying wave propagation phenomena can be purely unsteady or they can be periodic (because the imposed source term follows a time harmonic evolution). In this context, the overall objective of the research activities undertaken by the team is to develop numerical methods putting the emphasis on several features:

- Accuracy. The foreseen numerical methods should ideally rely on discretization techniques that best fit to the geometrical characteristics of the problems at hand. For this reason, the team focuses on methods working on unstructured, locally refined, even non-conforming, simplicial meshes. These methods should also be capable to accurately describe the underlying physical phenomena that may involve highly variable space and time scales. With reference to this characteristic, two main strategies are possible: adaptive local refinement/coarsening of the mesh (i.e h-adaptivity) and adaptive local variation of the interpolation order (i.e p-adaptivity). Ideally, these two strategies are combined leading to the so-called hp-adaptive methods.

- Numerical efficiency. The simulation of unsteady problems most often relies on explicit time integration schemes. Such schemes are constrained by a stability criteria linking the space and time discretization parameters that can be very restrictive when the underlying mesh is highly non-uniform (especially for locally refined meshes). For realistic 3D problems, this can represent a severe limitation with regards to the overall computing time. In order to improve this situation, one possible approach consists in resorting to an implicit time scheme in regions of the computational domain where the underlying mesh is refined while an explicit time scheme is applied to the remaining part of the domain. The resulting hybrid explicit-implicit time integration strategy raises several challenging questions concerning both the mathematical analysis (stability and accuracy, especially for what concern numerical dispersion), and the computer implementation on modern high performance systems (data structures, parallel computing aspects). A second, more classical approach is to devise a local time strategy in the context of a fully explicit time integration scheme. Stability and accuracy are still important challenges in this case.

  On the other hand, when considering time harmonic wave propagation problems, numerical efficiency is mainly linked to the solution of the system of algebraic equations resulting from the discretization in space of the underlying PDE model. Various strategies exist ranging from the more robust and efficient sparse direct solvers to the more flexible and cheaper (in terms of memory resources) iterative methods. Current trends tend to show that the ideal candidate will be a judicious mix of both approaches by relying on domain decomposition principles.

- Computational efficiency. Realistic 3D wave propagation problems lead to the processing of very large volumes of data. The latter results from two combined parameters: the size of the mesh i.e the number of mesh elements, and the number of degrees of freedom per mesh element which is itself linked to the degree of interpolation and to the number of physical variables (for systems of partial differential equations). Hence, numerical methods must be adapted to the characteristics of modern parallel computing platforms taking into account their hierarchical nature (e.g multiple processors and multiple core systems with complex cache and memory hierarchies). Besides, appropriate parallelization strategies need to be designed that combine SIMD and MIMD programming paradigms. Moreover, maximizing the effective floating point performances will require the design of numerical algorithms that can benefit from the optimized BLAS linear algebra kernels.

The discontinuous Galerkin method (DG) was introduced in 1973 by Reed and Hill to solve the neutron transport equation. From this time to the 90's a review on the DG methods would likely fit into one page. In the meantime, the finite volume approach has been widely adopted by computational fluid dynamics scientists and has now nearly supplanted classical finite difference and finite element methods in solving problems of non-linear convection. The success of the finite volume method is due to its ability to capture discontinuous solutions which may occur when solving non-linear equations or more simply, when convecting discontinuous initial data in the linear case. Let us first remark that DG methods share with finite volumes this property since a first order finite volume scheme can be viewed as a 0th order DG scheme. However a DG method may be also considered as a finite element one where the continuity constraint at an element interface is released. While it keeps almost all the advantages of the finite element method (large spectrum of applications, complex geometries, etc.), the DG method has other nice properties which explain the renewed interest it gains in various domains in scientific computing as witnessed by books or special issues of journals dedicated to this method [43]- [44]- [45]- [48]:

- It is naturally adapted to a high order approximation of the unknown field. Moreover, one may increase the degree of the approximation in the whole mesh as easily as for spectral methods but, with a DG method, this can also be done very locally. In most cases, the approximation relies on a polynomial interpolation method but the DG method also offers the flexibility of applying local approximation strategies that best fit to the intrinsic features of the modeled physical phenomena.

- When the discretization in space is coupled to an explicit time integration method, the DG method leads to a block diagonal mass matrix independently of the form of the local approximation (e.g the type of polynomial interpolation). This is a striking difference with classical, continuous finite element formulations. Moreover, the mass matrix is diagonal if an orthogonal basis is chosen.

- It easily handles complex meshes. The grid may be a classical conforming finite element mesh, a non-conforming one or even a hybrid mesh made of various elements (tetrahedra, prisms, hexahedra, etc.). The DG method has been proven to work well with highly locally refined meshes. This property makes the DG method more suitable to the design of a hp-adaptive solution strategy (i.e where the characteristic mesh size h and the interpolation degree p changes locally wherever it is needed).

- It is flexible with regards to the choice of the time stepping scheme. One may combine the DG spatial discretization with any global or local explicit time integration scheme, or even implicit, provided the resulting scheme is stable.

- It is naturally adapted to parallel computing. As long as an explicit time integration scheme is used, the DG method is easily parallelized. Moreover, the compact nature of DG discretization schemes is in favor of high computation to communication ratio especially when the interpolation order is increased.

As with standard finite element methods, a DG method relies on a variational formulation of the continuous problem at hand. However, due to the discontinuity of the global approximation, this variational formulation has to be defined at the element level. Then, a degree of freedom in the design of a DG method stems from the approximation of the boundary integral term resulting from the application of an integration by parts to the element-wise variational form. In the spirit of finite volume methods, the approximation of this boundary integral term calls for a numerical flux function which can be based on either a centered scheme or an upwind scheme, or a blending between these two schemes.

For the numerical solution of the time domain Maxwell equations, we have first proposed a non-dissipative high order DGTD (Discontinuous Galerkin Time Domain) method working on unstructured conforming simplicial meshes [14]-[3]. This DG method combines a central numerical flux function for the approximation of the integral term at an interface between two neighboring elements with a second order leap-frog time integration scheme. Moreover, the local approximation of the electromagnetic field relies on a nodal (Lagrange type) polynomial interpolation method. Recent achievements by the team deal with the extension of these methods towards non-conforming meshes and hp-adaptivity [12]-[13], their coupling with hybrid explicit/implicit time integration schemes in order to improve their efficiency in the context of locally refined meshes [6]. A high

order DG method has also been proposed for the numerical resolution of the elastodynamic equations modeling the propagation of seismic waves [5]-[11]. For the numerical treatment of the time harmonic Maxwell equations, we have studied similar DG methods [7]-[18] and more recently, HDG (Hybridized Discontinuous Galerkin) methods [22].

3.2. Domain decomposition methods

Domain Decomposition (DD) methods are flexible and powerful techniques for the parallel numerical solution of systems of PDEs. As clearly described in [51], they can be used as a process of distributing a computational domain among a set of interconnected processors or, for the coupling of different physical models applied in different regions of a computational domain (together with the numerical methods best adapted to each model) and, finally as a process of subdividing the solution of a large linear system resulting from the discretization of a system of PDEs into smaller problems whose solutions can be used to devise a parallel preconditioner or a parallel solver. In all cases, DD methods (1) rely on a partitioning of the computational domain into subdomains, (2) solve in parallel the local problems using a direct or iterative solver and, (3) call for an iterative procedure to collect the local solutions in order to get the global solution of the original problem. Subdomain solutions are connected by means of suitable transmission conditions at the artificial interfaces between the subdomains. The choice of these transmission conditions greatly influences the convergence rate of the DD method. One generally distinguish three kinds of DD methods:

- Overlapping methods use a decomposition of the computational domain in overlapping pieces. The so-called Schwarz method belongs to this class. Schwarz initially introduced this method for proving the existence of a solution to a Poisson problem. In the Schwarz method applied to the numerical resolution of elliptic PDEs, the transmission conditions at artificial subdomain boundaries are simple Dirichlet conditions. Depending on the way the solution procedure is performed, the iterative process is called a Schwarz multiplicative method (the subdomains are treated sequentially) or an additive method (the subdomains are treated in parallel).

- Non-overlapping methods are variants of the original Schwarz DD methods with no overlap between neighboring subdomains. In order to ensure convergence of the iterative process in this case, the transmission conditions are not trivial and are generally obtained through a detailed inspection of the mathematical properties of the underlying PDE or system of PDEs.

- Substructuring methods rely on a non-overlapping partition of the computational domain. They assume a separation of the problem unknowns in purely internal unknowns and interface ones. Then, the internal unknowns are eliminated thanks to a Schur complement technique yielding to the formulation of a problem of smaller size whose iterative resolution is generally easier. Nevertheless, each iteration of the interface solver requires the realization of a matrix/vector product with the Schur complement operator which in turn amounts to the concurrent solution of local subproblems.

Schwarz algorithms have enjoyed a second youth over the last decades, as parallel computers became more and more powerful and available. Fundamental convergence results for the classical Schwarz methods were derived for many partial differential equations, and can now be found in several books [51]- [50]- [53].

The research activities of the team on this topic aim at the formulation, analysis and evaluation of Schwarz type domain decomposition methods in conjunction with discontinuous Galerkin approximation methods on unstructured simplicial meshes for the solution of time domain and time harmonic wave propagation problems. Ongoing works in this direction are concerned with the design of non-overlapping Schwarz algorithms for the solution of the time harmonic Maxwell equations. A first achievement has been a Schwarz algorithm for the time harmonic Maxwell equations, where a first order absorbing condition is imposed at the interfaces between neighboring subdomains [9]. This interface condition is equivalent to a Dirichlet condition for characteristic variables associated to incoming waves. For this reason, it is often referred as a natural interface condition. Beside Schwarz algorithms based on natural interface conditions, the team also investigates algorithms that make use of more effective transmission conditions [10]. Recent contributions are concerned with the design and anlysis of such optimized Schwarz algorithm for the solution of the time harmonic Maxwell equations with non-zero conductivity [17].

3.3. High performance numerical computing

Beside basic research activities related to the design of numerical methods and resolution algorithms for the wave propagation models at hand, the team is also committed to demonstrate the benefits of the proposed numerical methodologies in the simulation of challenging three-dimensional problems pertaining to computational electromagnetics and computation geoseismics. For such applications, parallel computing is a mandatory path. Nowadays, modern parallel computers most often take the form of clusters of heterogeneous multiprocessor systems, combining multiple core CPUs with accelerator cards (e.g Graphical Processing Units - GPUs), with complex hierarchical distributed-shared memory systems. Developing numerical algorithms that efficiently exploit such high performance computing architectures raises several challenges, especially in the context of a massive parallelism. In this context, current efforts of the team are towards the exploitation of multiple levels of parallelism (computing systems combining CPUs and GPUs) through the study of hierarchical SPMD (Single Program Multiple Data) strategies for the parallelization of unstructured mesh based solvers.

OPALE Project-Team

3. Scientific Foundations

3.1. Functional and numerical analysis of PDE systems

Our common scientific background is the functional and numerical analysis of PDE systems, in particular with respect to nonlinear hyperbolic equations such as conservation laws of gas-dynamics.

Whereas the structure of weak solutions of the Euler equations has been thoroughly discussed in both the mathematical and fluid mechanics literature, in similar hyperbolic models, focus of new interest, such as those related to traffic, the situation is not so well established, except in one space dimension, and scalar equations. Thus, the study of such equations is one theme of emphasis of our research.

The well-developed domain of numerical methods for PDE systems, in particular finite volumes, constitute the sound background for PDE-constrained optimization.

3.2. Numerical optimization of PDE systems

Partial Differential Equations (PDEs), finite volumes/elements, geometrical optimization, optimum shape design, multi-point/multi-criterion/multi-disciplinary optimization, shape parameterization, gradient-based/evolutionary/hybrid optimizers, hierarchical physical/numerical models, Proper Orthogonal Decomposition (POD)

Optimization problems involving systems governed by PDEs, such as optimum shape design in aerodynamics or electromagnetics, are more and more complex in the industrial setting.

In certain situations, the major difficulty resides in the costly evaluation of a functional by means of a simulation, and the numerical method to be used must exploit at best the problem characteristics (regularity or smoothness, local convexity).

In many other cases, several criteria are to be optimized and some are non differentiable and/or non convex. A large set of parameters, sometimes of different types (boolean, integer, real or functional), are to be taken into account, as well as constraints of various types (physical and geometrical, in particular). Additionally, today's most interesting optimization pre-industrial projects are multi-disciplinary, and this complicates the mathematical, physical and numerical settings. Developing *robust optimizers* is therefore an essential objective to make progress in this area of scientific computing.

In the area of numerical optimization algorithms, the project aims at adapting classical optimization methods (simplex, gradient, quasi-Newton) when applicable to relevant engineering applications, as well as developing and testing less conventional approaches such as Evolutionary Strategies (ES), including Genetic or Particle-Swarm Algorithms, or hybrid schemes, in contexts where robustness is a very severe constraint.

In a different perspective, the heritage from the former project Sinus in Finite-Volumes (or -Elements) for nonlinear hyperbolic problems, leads us to examine cost-efficiency issues of large shape-optimization applications with an emphasis on the PDE approximation; of particular interest to us:

- best approximation and shape-parameterization,
- convergence acceleration (in particular by multi-level methods),
- model reduction (e.g. by *Proper Orthogonal Decomposition*),
- parallel and grid computing; etc.

3.3. Geometrical optimization

Jean-Paul Zolesio and Michel Delfour have developed, in particular in their book [6], a theoretical framework for for geometrical optimization and shape control in Sobolev spaces.

In preparation to the construction of sound numerical techniques, their contribution remains a fundamental building block for the functional analysis of shape optimization formulations.

3.4. Integration platforms

Developing grid, cloud and high-performance computing for complex applications is one of the priorities of the IST chapter in the 7th Framework Program of the European Community. One of the challenges of the 21st century in the computer science area lies in the integration of various expertise in complex application areas such as simulation and optimization in aeronautics, automotive and nuclear simulation. Indeed, the design of the reentry vehicle of a space shuttle calls for aerothermal, aerostructure and aerodynamics disciplines which all interact in hypersonic regime, together with electromagnetics. Further, efficient, reliable, and safe design of aircraft involve thermal flows analysis, consumption optimization, noise reduction for environmental safety, using for example aeroacoustics expertise.

The integration of such various disciplines requires powerful computing infrastructures and particular software coupling techniques. Simultaneously, advances in computer technology militate in favor of the use of massively parallel clusters including hundreds of thousands of processors connected by high-speed giga-bits/sec networks. This conjunction makes it possible for an unprecedented cross-fertilization of computational methods and computer science. New approaches including evolutionary algorithms, parameterization, multi-hierarchical decomposition lend themselves seamlessly to parallel implementations in such computing infrastructures. This opportunity is being dealt with by the OPALE project since its very beginning. A software integration platform has been designed by the OPALE project for the definition, configuration and deployment of multidisciplinary applications on a distributed heterogeneous infrastructure. Experiments conducted within European projects and industrial cooperations using CAST have led to significant performance results in complex aerodynamics optimization test-cases involving multi-elements airfoils and evolutionary algorithms, i.e. coupling genetic and hierarchical algorithms involving game strategies [70].

The main difficulty still remains however in the deployment and control of complex distributed applications by the end-users. Indeed, the deployment of the computing infrastructures and of the applications in such environments still requires specific expertise by computer science specialists. However, the users, which are experts in their particular application fields, e.g. aerodynamics, are not necessarily experts in distributed and grid computing. Being accustomed to Internet browsers, they want similar interfaces to interact with high-performance computing and problem-solving environments. A first approach to solve this problem is to define component-based infrastructures, e.g. the Corba Component Model, where the applications are considered as connection networks including various application codes. The advantage is here to implement a uniform approach for both the underlying infrastructure and the application modules. However, it still requires specific expertise not directly related to the application domains of each particular user. A second approach is to make use of web services, defined as application and support procedures to standardize access and invocation to remote support and application codes. This is usually considered as an extension of Web services to distributed infrastructures. A new approach, which is currently being explored by the OPALE project, is the design of a virtual computing environment able to hide the underlying high-performance-computing infrastructures to the users. The team is exploring the use of distributed workflows to define, monitor and control the execution of high-performance simulations on distributed clusters. The platform includes resilience, i.e., fault-tolerance features allowing for resource demanding and erroneous applications to be dynamically restarted safely, without user intervention.

3. Scientific Foundations

3.1. Automatic Differentiation

Participants: Laurent Hascoët, Valérie Pascual.

> **automatic differentiation** (AD) Automatic transformation of a program, that returns a new program that computes some derivatives of the given initial program, i.e. some combination of the partial derivatives of the program's outputs with respect to its inputs.
>
> **adjoint** Mathematical manipulation of the Partial Derivative Equations that define a problem, obtaining new differential equations that define the gradient of the original problem's solution.
>
> **checkpointing** General trade-off technique, used in adjoint-mode AD, that trades duplicate execution of a part of the program to save some memory space that was used to save intermediate results. Checkpointing a code fragment amounts to running this fragment without any storage of intermediate values, thus saving memory space. Later, when such an intermediate value is required, the fragment is run a second time to obtain the required values.

Automatic or Algorithmic Differentiation (AD) differentiates *programs*. An AD tool takes as input a source program P that, given a vector argument $X \in \mathbb{R}^n$, computes some vector result $Y = F(X) \in \mathbb{R}^m$. The AD tool generates a new source program P' that, given the argument X, computes some derivatives of F. The resulting P' reuses the control of P.

For any given control, P is equivalent to a sequence of instructions, which is identified with a composition of vector functions. Thus, if

$$
\begin{aligned}
P \quad &\text{is} \quad \{I_1; I_2; \cdots I_p; \}, \\
F \quad &= \quad f_p \circ f_{p-1} \circ \cdots \circ f_1,
\end{aligned}
\tag{1}
$$

where each f_k is the elementary function implemented by instruction I_k. AD applies the chain rule to obtain derivatives of F. Calling X_k the values of all variables after instruction I_k, i.e. $X_0 = X$ and $X_k = f_k(X_{k-1})$, the chain rule gives the Jacobian of F

$$
F'(X) = f'_p(X_{p-1}) \cdot f'_{p-1}(X_{p-2}) \cdot \; \cdots \; \cdot f'_1(X_0)
\tag{2}
$$

which can be mechanically written as a sequence of instructions I'_k. Combining the I'_k with the control of P yields P'. This can be generalized to higher level derivatives, Taylor series, etc.

In practice, the Jacobian $F'(X)$ is often too expensive to compute and store, but most applications only need projections of $F'(X)$ such as:

- **Sensitivities**, defined for a given direction $\dot{X}$ in the input space as:

$$
F'(X).\dot{X} = f'_p(X_{p-1}) \cdot f'_{p-1}(X_{p-2}) \cdot \; \cdots \; \cdot f'_1(X_0) \cdot \dot{X} \quad .
\tag{3}
$$

  Sensitivities are easily computed from right to left, interleaved with the original program instructions. This is the *tangent mode* of AD.

- **Adjoints**, defined for a given weighting $\overline{Y}$ of the outputs as:

$$F'^*(X).\overline{Y} = f_1'^*(X_0).f_2'^*(X_1). \ \cdots \ .f_{p-1}'^*(X_{p-2}).f_p'^*(X_{p-1}).\overline{Y} \quad . \tag{4}$$

Adjoints are most efficiently computed from right to left, because matrix×vector products are cheaper than matrix×matrix products. This is the *adjoint mode* of AD, most effective for optimization, data assimilation [33], adjoint problems [28], or inverse problems.

Adjoint-mode AD turns out to make a very efficient program, at least theoretically [30]. The computation time required for the gradient is only a small multiple of the run-time of P. It is independent from the number of parameters n. In contrast, computing the same gradient with the *tangent mode* would require running the tangent differentiated program n times.

However, the X_k are required in the *inverse* of their computation order. If the original program *overwrites* a part of X_k, the differentiated program must restore X_k before it is used by $f_{k+1}'^*(X_k)$. Therefore, the central research problem of adjoint-mode AD is to make the X_k available in reverse order at the cheapest cost, using strategies that combine storage, repeated forward computation from available previous values, or even inverted computation from available later values.

Another research issue is to make the AD model cope with the constant evolution of modern language constructs. From the old days of Fortran77, novelties include pointers and dynamic allocation, modularity, structured data types, objects, vectorial notation and parallel communication. We regularly extend our models and tools to handle these new constructs.

3.2. Static Analysis and Transformation of programs

Participants: Laurent Hascoët, Valérie Pascual.

> **abstract syntax tree**   Tree representation of a computer program, that keeps only the semantically significant information and abstracts away syntactic sugar such as indentation, parentheses, or separators.
>
> **control flow graph**   Representation of a procedure body as a directed graph, whose nodes, known as basic blocks, contain each a list of instructions to be executed in sequence, and whose arcs represent all possible control jumps that can occur at run-time.
>
> **abstract interpretation**   Model that describes program static analysis as a special sort of execution, in which all branches of control switches are taken simultaneously, and where computed values are replaced by abstract values from a given *semantic domain*. Each particular analysis gives birth to a specific semantic domain.
>
> **data flow analysis**   Program analysis that studies how a given property of variables evolves with execution of the program. Data Flow analysis is static, therefore studying all possible run-time behaviors and making conservative approximations. A typical data-flow analysis is to detect whether a variable is initialized or not, at any location in the source program.
>
> **data dependence analysis**   Program analysis that studies the itinerary of values during program execution, from the place where a value is generated to the places where it is used, and finally to the place where it is overwritten. The collection of all these itineraries is often stored as a *data dependence graph*, and data flow analysis most often rely on this graph.
>
> **data dependence graph**   Directed graph that relates accesses to program variables, from the write access that defines a new value to the read accesses that use this value, and conversely from the read accesses to the write access that overwrites this value. Dependences express a partial order between operations, that must be preserved to preserve the program's result.

The most obvious example of a program transformation tool is certainly a compiler. Other examples are program translators, that go from one language or formalism to another, or optimizers, that transform a program to make it run better. AD is just one such transformation. These tools use sophisticated analysis [20] to improve the quality of the produced code. These tools share their technological basis. More importantly, there are common mathematical models to specify and analyze them.

An important principle is *abstraction*: the core of a compiler should not bother about syntactic details of the compiled program. The optimization and code generation phases must be independent from the particular input programming language. This is generally achieved using language-specific *front-ends* and *back-ends*. Abstraction can go further: the internal representation becomes more language independent, and semantic constructs can be unified. Analysis can then concentrate on the semantics of a small set of constructs. We advocate an internal representation composed of three levels.

- At the top level is the *call graph*, whose nodes are modules and procedures. Arrows relate nodes that call or import one another. Recursion leads to cycles.
- At the middle level is the *flow graph*, one per procedure. It captures the control flow between atomic instructions.
- At the lowest level are abstract *syntax trees* for the individual atomic instructions. Semantic transformations can benefit from the representation of expressions as directed acyclic graphs, sharing common sub-expressions.

To each level belong symbol tables, nested to capture scoping.

Static program analysis can be defined on this internal representation, which is largely language independent. The simplest analyses on trees can be specified with inference rules [23], [31], [21]. But many analyses are more complex, and better defined on graphs than on trees. This is the case for *data-flow analyses*, that look for run-time properties of variables. Since flow graphs are cyclic, these global analyses generally require an iterative resolution. *Data flow equations* is a practical formalism to describe data-flow analyses. Another formalism is described in [24], which is more precise because it can distinguish separate *instances* of instructions. However it is still based on trees, and its cost forbids application to large codes. *Abstract Interpretation* [25] is a theoretical framework to study complexity and termination of these analyses.

Data flow analyses must be carefully designed to avoid or control combinatorial explosion. At the call graph level, they can run bottom-up or top-down, and they yield more accurate results when they take into account the different call sites of each procedure, which is called *context sensitivity*. At the flow graph level, they can run forwards or backwards, and yield more accurate results when they take into account only the possible execution flows resulting from possible control, which is called *flow sensitivity*.

Even then, data flow analyses are limited, because they are static and thus have very little knowledge of actual run-time values. In addition to the very theoretical limit of *undecidability*, there are practical limitations to how much information one can infer from programs that use arrays [37], [26] or pointers. In general, conservative *over-approximations* are always made that lead to derivative code that is less efficient than possibly achievable.

3.3. Automatic Differentiation and Scientific Computing

Participants: Alain Dervieux, Laurent Hascoët, Bruno Koobus.

linearization   In Scientific Computing, the mathematical model often consists of Partial Derivative Equations, that are discretized and then solved by a computer program. Linearization of these equations, or alternatively linearization of the computer program, predict the behavior of the model when small perturbations are applied. This is useful when the perturbations are effectively small, as in acoustics, or when one wants the sensitivity of the system with respect to one parameter, as in optimization.

adjoint state   Consider a system of Partial Derivative Equations that define some characteristics of a system with respect to some input parameters. Consider one particular scalar characteristic. Its sensitivity, (or gradient) with respect to the input parameters can be defined as the solution of "adjoint" equations, deduced from the original equations through linearization and transposition. The solution of the adjoint equations is known as the adjoint state.

Scientific Computing provides reliable simulations of complex systems. For example it is possible to simulate the 3D air flow around a plane that captures the physical phenomena of shocks and turbulence. Next comes optimization, one degree higher in complexity because it repeatedly simulates and applies optimization steps until an optimum is reached. We focus on gradient-based optimization.

We investigate several approaches to obtain the gradient, between two extremes:

- One can write an *adjoint system* of mathematical equations, then discretize it and program it by hand. This is mathematically sound [28], but very costly in development time. It also does not produce an exact gradient of the discrete function, and this can be a problem if using optimization methods based on descent directions.

- One can apply adjoint-mode AD (*cf*3.1) on the program that discretizes and solves the direct system. This gives in fact the adjoint of the discrete function computed by the program. Theoretical results [27] guarantee convergence of these derivatives when the direct program converges. This approach is highly mechanizable, but leads to massive use of storage and may require code transformation by hand [32], [35] to reduce memory usage.

If for instance the model is steady, one tradeoff can use the iterated states in the direct order [29]. Another tradeoff can use only the fully converged final state. Since tradeoff approaches can be error-prone, we advocate incorporating them into the AD model and into the AD tools.

<h2 style="text-align:center;color:red;">TOSCA Project-Team</h2>

3. Scientific Foundations

3.1. Scientific Foundations

Most often physicists, economists, biologists, engineers need a stochastic model because they cannot describe the physical, economical, biological, etc., experiment under consideration with deterministic systems, either because of its complexity and/or its dimension or because precise measurements are impossible. Then they abandon trying to get the exact description of the state of the system at future times given its initial conditions, and try instead to get a statistical description of the evolution of the system. For example, they desire to compute occurrence probabilities for critical events such as the overstepping of a given thresholds by financial losses or neuronal electrical potentials, or to compute the mean value of the time of occurrence of interesting events such as the fragmentation to a very small size of a large proportion of a given population of particles. By nature such problems lead to complex modelling issues: one has to choose appropriate stochastic models, which require a thorough knowledge of their qualitative properties, and then one has to calibrate them, which requires specific statistical methods to face the lack of data or the inaccuracy of these data. In addition, having chosen a family of models and computed the desired statistics, one has to evaluate the sensitivity of the results to the unavoidable model specifications. The TOSCA team, in collaboration with specialists of the relevant fields, develops theoretical studies of stochastic models, calibration procedures, and sensitivity analysis methods.

In view of the complexity of the experiments, and thus of the stochastic models, one cannot expect to use closed form solutions of simple equations in order to compute the desired statistics. Often one even has no other representation than the probabilistic definition (e.g., this is the case when one is interested in the quantiles of the probability law of the possible losses of financial portfolios). Consequently the practitioners need Monte Carlo methods combined with simulations of stochastic models. As the models cannot be simulated exactly, they also need approximation methods which can be efficiently used on computers. The TOSCA team develops mathematical studies and numerical experiments in order to determine the global accuracy and the global efficiency of such algorithms.

The simulation of stochastic processes is not motivated by stochastic models only. The stochastic differential calculus allows one to represent solutions of certain deterministic partial differential equations in terms of probability distributions of functionals of appropriate stochastic processes. For example, elliptic and parabolic linear equations are related to classical stochastic differential equations, whereas nonlinear equations such as the Burgers and the Navier–Stokes equations are related to McKean stochastic differential equations describing the asymptotic behavior of stochastic particle systems. In view of such probabilistic representations one can get numerical approximations by using discretization methods of the stochastic differential systems under consideration. These methods may be more efficient than deterministic methods when the space dimension of the PDE is large or when the viscosity is small. The TOSCA team develops new probabilistic representations in order to propose probabilistic numerical methods for equations such as conservation law equations, kinetic equations, and nonlinear Fokker–Planck equations.

ABS Project-Team

3. Scientific Foundations

3.1. Introduction

The research conducted by ABS focuses on two main directions in Computational Structural Biology (CSB), each such direction calling for specific algorithmic developments. These directions are:
- Modeling interfaces and contacts,
- Modeling the flexibility of macro-molecules.

3.2. Modeling Interfaces and Contacts

Docking, interfaces, protein complexes, structural alphabets, scoring functions, Voronoi diagrams, arrangements of balls

Problems addressed. The Protein Data Bank, http://www.rcsb.org/pdb, contains the structural data which have been resolved experimentally. Most of the entries of the PDB feature isolated proteins [1], the remaining ones being protein - protein or protein - drug complexes. These structures feature what Nature does —up to the bias imposed by the experimental conditions inherent to structure elucidation, and are of special interest to investigate non-covalent contacts in biological complexes. More precisely, given two proteins defining a complex, interface atoms are defined as the atoms of one protein *interacting* with atoms of the second one. Understanding the structure of interfaces is central to understand biological complexes and thus the function of biological molecules [46]. Yet, in spite of almost three decades of investigations, the basic principles guiding the formation of interfaces and accounting for its stability are unknown [49]. Current investigations follow two routes. From the experimental perspective [31], directed mutagenesis enables one to quantify the energetic importance of residues, important residues being termed *hot* residues. Such studies recently evidenced the *modular* architecture of interfaces [43]. From the modeling perspective, the main issue consists of guessing the hot residues from sequence and/or structural informations [38].

The description of interfaces is also of special interest to improve *scoring functions*. By scoring function, two things are meant: either a function which assigns to a complex a quantity homogeneous to a free energy change [2], or a function stating that a complex is more stable than another one, in which case the value returned is a score and not an energy. Borrowing to statistical mechanics [23], the usual way to design scoring functions is to mimic the so-called potentials of mean force. To put it briefly, one reverts Boltzmann's law, that is, denoting $p_i(r)$ the probability of two atoms –defining type i– to be located at distance r, the (free) energy assigned to the pair is computed as $E_i(r) = -kT \log p_i(r)$. Estimating from the PDB one function $p_i(r)$ for each type of pair of atoms, the energy of a complex is computed as the sum of the energies of the pairs located within a distance threshold [47], [34]. To compare the energy thus obtained to a reference state, one may compute $E = \sum_i p_i \log p_i / q_i$, with p_i the observed frequencies, and q_i the frequencies stemming from an a priori model [39]. In doing so, the energy defined is nothing but the Kullback-Leibler divergence between the distributions $\{p_i\}$ and $\{q_i\}$.

Methodological developments. Describing interfaces poses problems in two settings: static and dynamic.

---

[1]For structures resolved by crystallography, the PDB contains the asymmetric unit of the crystal. Determining the biological unit from the asymmetric unit is a problem in itself.

[2]The Gibbs free energy of a system is defined by $G = H - TS$, with $H = U + PV$. G is minimum at an equilibrium, and differences in G drive chemical reactions.

In the static setting, one seeks the minimalist geometric model providing a relevant bio-physical signal. A first step in doing so consists of identifying interface atoms, so as to relate the geometry and the bio-chemistry at the interface level [10]. To elaborate at the atomic level, one seeks a structural alphabet encoding the spatial structure of proteins. At the side-chain and backbone level, an example of such alphabet is that of [24]. At the atomic level and in spite of recent observations on the local structure of the neighborhood of a given atom [48], no such alphabet is known. Specific important local conformations are known, though. One of them is the so-called dehydron structure, which is an under-desolvated hydrogen bond —a property that can be directly inferred from the spatial configuration of the C_α carbons surrounding a hydrogen bond [30].

A structural alphabet at the atomic level may be seen as an alphabet featuring for an atom of a given type all the conformations this atom may engage into, depending on its neighbors. One way to tackle this problem consists of extending the notions of molecular surfaces used so far, so as to encode multi-body relations between an atom and its neighbors [8]. In order to derive such alphabets, the following two strategies are obvious. On one hand, one may use an encoding of neighborhoods based on geometric constructions such as Voronoi diagrams (affine or curved) or arrangements of balls. On the other hand, one may resort to clustering strategies in higher dimensional spaces, as the p neighbors of a given atom are represented by $3p - 6$ degrees of freedom —the neighborhood being invariant upon rigid motions.

In the dynamic setting, one wishes to understand whether selected (hot) residues exhibit specific dynamic properties, so as to serve as anchors in a binding process [42]. More generally, any significant observation raised in the static setting deserves investigations in the dynamic setting, so as to assess its stability. Such questions are also related to the problem of correlated motions, which we discuss next.

3.3. Modeling Macro-molecular Assemblies

Macro-molecular assembly, reconstruction by data integration, proteomics, modeling with uncertainties, curved Voronoi diagrams, topological persistence.

3.3.1. *Reconstruction by data integration*

Large protein assemblies such as the Nuclear Pore Complex (NPC), chaperonin cavities, the proteasome or ATP synthases, to name a few, are key to numerous biological functions. To improve our understanding of these functions, one would ideally like to build and animate atomic models of these molecular machines. However, this task is especially tough, due to their size and their plasticity, but also due to the flexibility of the proteins involved. In a sense, the modeling challenges arising in this context are different from those faced for binary docking, and also from those encountered for intermediate size complexes which are often amenable to a processing mixing (cryo-EM) image analysis and classical docking. To face these new challenges, an emerging paradigm is that of reconstruction by data integration [22]. In a nutshell, the strategy is reminiscent from NMR and consists of mixing experimental data from a variety of sources, so as to find out the model(s) best complying with the data. This strategy has been in particular used to propose plausible models of the Nuclear Pore Complex [21], the largest assembly known to date in the eukaryotic cell, and consisting of 456 protein *instances* of 30 *types*.

3.3.2. *Modeling with uncertainties and model assessment*

Reconstruction by data integration requires three ingredients. First, a parametrized model must be adopted, typically a collection of balls to model a protein with pseudo-atoms. Second, as in NMR, a functional measuring the agreement between a model and the data must be chosen. In [20], this functional is based upon *restraints*, namely penalties associated to the experimental data. Third, an optimization scheme must be selected. The design of restraints is notoriously challenging, due to the ambiguous nature and/or the noise level of the data. For example, Tandem Affinity Purification (TAP) gives access to a *pullout* i.e. a list of protein types which are known to interact with one tagged protein type, but no information on the number of complexes or on the stoichiometry of proteins types within a complex is provided. In cryo-EM, the envelope enclosing an assembly is often imprecisely defined, in particular in regions of low density. For immuno-EM labelling experiments, positional uncertainties arise from the microscope resolution.

These uncertainties coupled with the complexity of the functional being optimized, which in general is non convex, have two consequences. First, it is impossible to single out a unique reconstruction, and a set of plausible reconstructions must be considered. As an example, 1000 plausible models of the NPC were reported in [20]. Interestingly, averaging the positions of all balls of a particular protein type across these models resulted in 30 so-called *probability density maps*, each such map encoding the probability of presence of a particular protein type at a particular location in the NPC. Second, the assessment of all models (individual and averaged) is non trivial. In particular, the lack of straightforward statistical analysis of the individual models and the absence of assessment for the averaged models are detrimental to the mechanistic exploitation of the reconstruction results. At this stage, such models therefore remain qualitative.

3.3.3. *Methodological developments*

As outlined by the previous discussion, a number of methodological developments are called for. On the experimental side, the problem of fostering the interpretation of data is under scrutiny. Of particular interest is the disambiguation of proteomics signals (TAP data, mass spectrometry data), and that of density maps coming from electron microscopy. As for modeling, two classes of developments are particularly stimulating. The first one is concerned with the design of algorithms performing reconstruction by data integration. The second one encompasses assessment tools, in order to single out the reconstructions which best comply with the experimental data.

3.4. Modeling the Flexibility of Macro-molecules

Folding, docking, energy landscapes, induced fit, molecular dynamics, conformers, conformer ensembles, point clouds, reconstruction, shape learning, Morse theory
Problems addressed. Proteins in vivo vibrate at various frequencies: high frequencies correspond to small amplitude deformations of chemical bonds, while low frequencies characterize more global deformations. This flexibility contributes to the entropy thus the `free energy` of the system *protein - solvent*. From the experimental standpoint, NMR studies and Molecular Dynamics simulations generate ensembles of conformations, called `conformers`. Of particular interest while investigating flexibility is the notion of correlated motion. Intuitively, when a protein is folded, all atomic movements must be correlated, a constraint which gets alleviated when the protein unfolds since the steric constraints get relaxed [3]. Understanding correlations is of special interest to predict the folding pathway that leads a protein towards its native state. A similar discussion holds for the case of partners within a complex, for example in the third step of the *diffusion - conformer selection - induced fit* complex formation model.

Parameterizing these correlated motions, describing the corresponding energy landscapes, as well as handling collections of conformations pose challenging algorithmic problems.

Methodological developments. At the side-chain level, the question of improving rotamer libraries is still of interest [29]. This question is essentially a clustering problem in the parameter space describing the side-chains conformations.

At the atomic level, flexibility is essentially investigated resorting to methods based on a classical potential energy (molecular dynamics), and (inverse) kinematics. A molecular dynamics simulation provides a point cloud sampling the conformational landscape of the molecular system investigated, as each step in the simulation corresponds to one point in the parameter space describing the system (the conformational space) [45]. The standard methodology to analyze such a point cloud consists of resorting to normal modes. Recently, though, more elaborate methods resorting to more local analysis [41], to Morse theory [36] and to analysis of meta-stable states of time series [37] have been proposed.

---

[3]Assuming local forces are prominent, which in turn subsumes electrostatic interactions are not prominent.

Given a sampling on an energy landscape, a number of fundamental issues actually arise: how does the point cloud describe the topography of the energy landscape (a question reminiscent from Morse theory)? Can one infer the effective number of degrees of freedom of the system over the simulation, and is this number varying? Answers to these questions would be of major interest to refine our understanding of folding and docking, with applications to the prediction of structural properties. It should be noted in passing that such questions are probably related to modeling phase transitions in statistical physics where geometric and topological methods are being used [40].

From an algorithmic standpoint, such questions are reminiscent of *shape learning*. Given a collection of samples on an (unknown) *model*, *learning* consists of guessing the model from the samples —the result of this process may be called the *reconstruction*. In doing so, two types of guarantees are sought: topologically speaking, the reconstruction and the model should (ideally!) be isotopic; geometrically speaking, their Hausdorff distance should be small. Motivated by applications in Computer Aided Geometric Design, surface reconstruction triggered a major activity in the Computational Geometry community over the past ten years [6]. Aside from applications, reconstruction raises a number of deep issues: the study of distance functions to the model and to the samples, and their comparison [25]; the study of Morse-like constructions stemming from distance functions to points [33]; the analysis of topological invariants of the model and the samples, and their comparison [26], [27].

Last but not least, gaining insight on such questions would also help to effectively select a reduced set of conformations best representing a larger number of conformations. This selection problem is indeed faced by flexible docking algorithms that need to maintain and/or update collections of conformers for the second stage of the *diffusion - conformer selection - induced fit* complex formation model.

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

3. Scientific Foundations

3.1. Introduction

Tremendous progress has been made in the automated analysis of biomedical images during the past two decades [91]. Readers who are neophyte to the field of medical imaging will find an interesting presentation of acquisition techniques of the main medical imaging modalities in [82], [80]. Regarding the target applications, a good review of the state of the art can be found in the book *Computer Integrated Surgery* [78], in N. Ayache's article [86] and in the more recent syntheses [87] [91]. The scientific journals *Medical Image Analysis* [73], *Transactions on Medical Imaging* [79], and *Computer Assisted Surgery* [81] are also good reference material. One can have a good vision of the state of the art with the proceedings of the most recent conferences MICCAI'2010 (Medical Image Computing and Computer Assisted Intervention) [76], [77] or ISBI'2010 (Int. Symp. on Biomedical Imaging) [75].

For instance, for rigid parts of the body like the head, it is now possible to fuse in a completely automated manner images of the same patient taken from different imaging modalities (e.g. anatomical and functional), or to track the evolution of a pathology through the automated registration and comparison of a series of images taken at distant time instants [92], [106]. It is also possible to obtain from a Magnetic Resonance Image (MRI) of the head a reasonable segmentation into skull tissues, white matter, grey matter, and cerebro-spinal fluid [109], or to measure some functional properties of the heart from dynamic sequences of Magnetic Resonance [85], Ultrasound or Nuclear Medicine images [93].

Despite these advances and successes, one can notice that statistical models of the anatomy are still very crude, resulting in poor registration results in deformable regions of the body, or between different subjects. If some algorithms exploit the physical modeling of the image acquisition process, only a few actually model the physical or even physiological properties of the human body itself. Coupling biomedical image analysis with anatomical and physiological models of the human body could not only provide a better comprehension of the observed images and signals, but also more efficient tools to detect anomalies, predict evolutions, simulate and assess therapies.

3.2. Medical Image Analysis

The quality of biomedical images tends to improve constantly (better spatial and temporal resolution, better signal to noise ratio). Not only the images are multidimensional (3 spatial coordinates and possibly one temporal dimension), but medical protocols tend to include multi-sequence (or multi-parametric) [1] and multi-modal images [2] for each single patient.

---

[1]Multisequence (or multiparametric) imaging consists in acquiring several images of a given patient with the same imaging modality (e.g. MRI, CT, US, SPECT, etc.) but with varying acquisition parameters. For instance, using Magnetic Resonance Imaging (MRI), patients followed for multiple sclerosis may undergo every six months a 3-D multisequence MR acquisition protocol with different pulse sequences (called T1, T2, PD, Flair etc): by varying some parameters of the pulse sequences (e.g Echo Time and Repetition Time), images of the same regions are produced with quite different contrasts depending on the nature and function of the observed structures. In addition, one of the acquisition (T1) can be combined with the injection of a contrast product (typically Gadolinium) to reveal vessels and some pathologies. Diffusion tensor images (DTI) can be acquired to measure the self diffusion of protons in every voxel, allowing to measure for instance the direction of white matter fibers in the brain (same principle can be used to measure the direction of muscular fibers in the heart). Functional MR images of the brain can be acquired by exploiting the so-called Bold Effect (Blood Oxygen Level Dependency): slightly higher blood flow in active regions creates subtle higher T2* signal which can be detected with sophisticated image processing techniques.

[2]Multimodal acquisition consists in acquiring on the same patient images from different modalities, in order to exploit their complementary nature. For instance CT and MR may provide information on the anatomy (CT providing contrast between bones and soft tissues, MR providing contrast within soft tissues of different nature) while SPECT and PET images may provide functional information by measuring a local level of metabolic activity.

Despite remarkable efforts and advances during the past twenty years, the central problems of segmentation and registration have not been solved in the general case. It is our objective in the short term to work on specific versions of these problems, taking into account as much *a priori* information as possible on the underlying anatomy and pathology at hand. It is also our objective to include more knowledge on the physics of image acquisition and observed tissues, as well as on the biological processes involved. Therefore the research activities mentioned in this section will incorporate the advances made in Computational Anatomy and Computational Physiology as described in sections 3.4 and 3.5 .

We plan to pursue our efforts on the following problems:

1. multi-dimensional, multi-sequence and multi-modal image segmentation,
2. Image Registration/Fusion,

3.3. Biological Image Analysis

In biology, a huge number of images of living systems are produced every day to study the basic mechanisms of life and pathologies. If some bio-imaging *principles* are the same as the ones used for medical applications (e.g. MR, CT, US, PET or SPECT), the bio-imaging *devices* are usually customized to produce images of higher resolution [3] for the observation of small animals (typically rodents). In addition, Optical Imaging (OI) techniques and biophotonics are developing very fast. This includes traditional or Confocal Microscopy (CM), multi-photon confocal microscopy, Optical Coherent Tomography (OCT), near-infrared imaging, diffuse optical imaging, phased array imaging, etc. A very new and promising development concerns micro-endoscopy, which allows cellular imaging at the end of a very small optical fiber [98].

Most of these imaging techniques can be used for *Molecular Imaging*, an activity aiming at the *in vivo* characterization and measurement of biological processes at cellular and molecular levels. With optical techniques, molecular imaging makes an extensive use of the fluorescent properties of certain molecules (in particular proteins, e.g. GFP [4]) for imaging of gene expression *in vivo*. With other modalities (like PET, SPECT, MR, CT and even US), molecular imaging can use specific contrast agents or radioactive molecules. For clinical applications, the ultimate goal of molecular imaging is to find the ways to probe much earlier the molecular anomalies that are the basis of a disease rather than to image only its end effects [110].

Some of the recent advances made in Medical Image Analysis could be directly applied (or easily adapted) to Biological Image Analysis. However, the specific nature of biological images (higher resolution, different anatomy and functions, different contrast agents, etc.), requires specific image analysis methods (one can refer to the recent tutorial [103] and to the Mouse Brain Atlas Project [84]). This is particularly true when dealing with *in vivo* microscopic images of cells and vessels.

Our research efforts will be focused to the following generic problems applied to *in vivo* microscopic images:

1. quantitative analysis of microscopic images,
2. detection and quantification of variations in temporal sequences,
3. construction of multiscale representations (from micro to macro).

3.4. Computational Anatomy

The objective of Computational Anatomy (CA) is the modeling and analysis of biological variability of the human anatomy. Typical applications cover the simulation of average anatomies and normal variations, the discovery of structural differences between healthy and diseased populations, and the detection and classification of pathologies from structural anomalies [5].

---

[3]This is the case with micro-MRI, Micro-CT, Micro-US devices, and to a less extent with Micro-SPECT and Micro-PET devices.
[4]Green Fluorescent Protein.
[5]The NIH has lauched the Alzheimer's Disease Neuroimaging Initiative (60 million USD), a multi-center MRI study of 800 patients who will be followed during several years. The objective will be to establish new surrogate end-points from the automated analysis of temporal sequences. This is a challenging objective for researchers in Computational Anatomy. The data will be made available to qualified research groups involved or not in the study.

Studying the variability of biological shapes is an old problem (cf. the remarkable book "On Shape and Growth" by D'Arcy Thompson [108]). Significant efforts have been made since that time to develop a theory for statistical shape analysis (one can refer to [90] for a good synthesis, and to the special issue of Neuroimage [107] for recent developments). Despite all these efforts, there is a number of challenging mathematical issues which remain largely unsolved in general. A particular issue is the computation of statistics on manifolds which can be of infinite dimension (e.g the group of diffeomorphisms).

There is a classical stratification of the problems into the following 3 levels [100]: 1) construction from medical images of anatomical manifolds of points, curves, surfaces and volumes; 2) assignment of a point to point correspondence between these manifolds using a specified class of transformations (e.g. rigid, affine, diffeomorphism); 3) generation of probability laws of anatomical variation from these correspondences.

We plan to focus our efforts to the following problems:

1. Statistics on anatomical manifolds,
2. Propagation of variability from anatomical manifolds,
3. Linking anatomical variability to image analysis algorithms,
4. Grid-Computing Strategies to exploit large databases.

3.5. Computational Physiology

The objective of Computational Physiology (CP) is to provide models of the major functions of the human body and numerical methods to simulate them. The main applications are in medicine and biology, where CP can be used for instance to better understand the basic processes leading to the apparition of a pathology, to model its probable evolution and to plan, simulate, and monitor its therapy.

Quite advanced models have already been proposed to study at the molecular, cellular and organic level a number of physiological systems (see for instance [102], [97], [88], [104], [94]). While these models and new ones need to be developed, refined or validated, a grand challenge that we want to address in this project is the automatic adaptation of the model to a given patient by confronting the model with the available biomedical images and signals and possibly also from some additional information (e.g. genetic). Building such *patient-specific models* is an ambitious goal which requires the choice or construction of models with a complexity adapted to the resolution of the accessible measurements (e.g. [105], [101]) and the development of new data assimilation methods coping with massive numbers of measurements and unknowns.

There is a hierarchy of modeling levels for CP models of the human body [89]:

- the first level is mainly geometrical, and addresses the construction of a digital description of the anatomy [83], essentially acquired from medical imagery;
- the second level is physical, involving mainly the biomechanical modeling of various tissues, organs, vessels, muscles or bone structures [95];
- the third level is physiological, involving a modeling of the functions of the major biological systems [96] (e.g. cardiovascular, respiratory, digestive, central or peripheral nervous, muscular, reproductive, hormonal, etc.) or some pathological metabolism (e.g. evolution of cancerous or inflammatory lesions, formation of vessel stenoses, etc.);
- a fourth level would be cognitive, modeling the higher functions of the human brain [74].

These different levels of modeling are closely related to each other, and several physiological systems may interact together (e.g. the cardiopulmonary interaction [99]). The choice of the resolution at which each level is described is important, and may vary from microscopic to macroscopic, ideally through multiscale descriptions.

Building this complete hierarchy of models is necessary to evolve from a *Visible Human* project (essentially first level of modeling) to a much more ambitious *Physiological Human project* (see [96], [97]). We will not address all the issues raised by this ambitious project, but instead focus on topics detailed below. Among them, our objective is to identify some common methods for the resolution of the large inverse problems raised by the coupling of physiological models to biological images for the construction of patient-specific models (e.g. specific variational or sequential methods (EKF), dedicated particle filters, etc.). We also plan to develop a specific expertise on the extraction of geometrical meshes from medical images for their further use in simulation procedures. Finally, computational models can be used for specific image analysis problems studied in section 3.2 (e.g. segmentation, registration, tracking, etc.). Application domains include

1. Surgery Simulation,

2. Cardiac Imaging,

3. Brain tumors, neo-angiogenesis, wound healing processes, ovocyte regulation, ...

3.6. Clinical and Biological Validation

If the objective of many of the research activities of the project is the discovery of original methods and algorithms with a demonstration of feasibility on a limited number of representative examples (i.e. proofs of concept) and publications in high quality scientific journals, we believe that it is important that a reasonable number of studies include a much more significant validation effort. As the BioMedical Image Analysis discipline becomes more mature, this is a necessary condition to see new ideas transformed into clinical tools and/or industrial products. It is also often the occasion to get access to larger databases of images and signals which in turn participate to the stimulation of new ideas and concepts.

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

3. Scientific Foundations

3.1. Computational Diffusion MRI

Diffusion MRI (dMRI) provides a non-invasive way of estimating in-vivo CNS fiber structures using the average random thermal movement (diffusion) of water molecules as a probe. It's a recent field of research with a history of roughly three decades. It was introduced in the mid 80's by Le Bihan et al [59], Merboldt et al [64] and Taylor et al [70]. As of today, it is the unique non-invasive technique capable of describing the neural connectivity in vivo by quantifying the anisotropic diffusion of water molecules in biological tissues. The great success of dMRI comes from its ability to accurately describe the geometry of the underlying microstructure and probe the structure of the biological tissue at scales much smaller than the imaging resolution.

The diffusion of water molecules is Brownian in an isotropic medium and under normal unhindered conditions, but in fibrous structure such as white matter, the diffusion is very often directionally biased or anisotropic and water molecules tend to diffuse along fibers. For example, a molecule inside the axon of a neuron has a low probability to cross a myelin membrane. Therefore the molecule will move principally along the axis of the neural fiber. Conversely if we know that molecules locally diffuse principally in one direction, we can make the assumption that this corresponds to a set of fibers.

Diffusion Tensor Imaging

Shortly after the first acquisitions of diffusion-weighted images (DWI) were made in vivo [65], [66], Basser et al [45], [44] proposed the rigorous formalism of the second order Diffusion Tensor Imaging model (DTI). DTI describes the three-dimensional (3D) nature of anisotropy in tissues by assuming that the average diffusion of water molecules follows a Gaussian distribution. It encapsulates the diffusion properties of water molecules in biological tissues (inside a typical 1-3 mm^3 sized voxel) as an effective self-diffusion tensor given by a 3×3 symmetric positive definite tensor $\mathbf{D}$ [45], [44]. Diffusion tensor imaging (DTI) thus produces a three-dimensional image containing, at each voxel, the estimated tensor $\mathbf{D}$. This requires the acquisition of at least six Diffusion Weighted Images (DWI) S_k in several non-coplanar encoding directions as well as an unweighted image S_0. Because of the signal attenuation, the image noise will affect the measurements and it is therefore important to take into account the nature and the strength of this noise in all the pre-processing steps. From the diffusion tensor $\mathbf{D}$, a neural fiber direction can be inferred from the tensor's main eigenvector while various diffusion anisotropy measures, such as the Fractional Anisotropy (FA), can be computed using the associated eigenvalues to quantify anisotropy, thus describing the inequality of diffusion values among particular directions.

DTI has now proved to be extremely useful to study the normal and pathological human brain [60], [51]. It has led to many applications in clinical diagnosis of neurological diseases and disorder, neurosciences applications in assessing connectivity of different brain regions, and more recently, therapeutic applications, primarily in neurosurgical planning. An important and very successful application of diffusion MRI has been brain ischemia, following the discovery that water diffusion drops immediately after the onset of an ischemic event, when brain cells undergo swelling through cytotoxic edema.

The increasing clinical importance of diffusion imaging has drived our interest to develop new processing tools for Diffusion MRI. Because of the complexity of the data, this imaging modality raises a large amount of mathematical and computational challenges. We have therefore started to develop original and efficient algorithms relying on Riemannian geometry, differential geometry, partial differential equations and front propagation techniques to correctly and efficiently estimate, regularize, segment and process Diffusion Tensor MRI (DT-MRI) (see [62], [8] and [61]).

High Angular Resolution Diffusion Imaging

In DTI, the Gaussian assumption over-simplifies the diffusion of water molecules. While it is adequate for voxels in which there is only a single fiber orientation (or none), it breaks for voxels in which there are more complex internal structures. This is an important limitation, since resolution of DTI acquisition is between $1mm^3$ and $3mm^3$ while the physical diameter of fibers can be between $1\mu m$ and 30 μm [68], [46]. Research groups currently agree that there is complex fiber architecture in most fiber regions of the brain [67]. In fact, it is currently thought that between one third to two thirds of imaging voxels in the human brain white matter contain multiple fiber bundle crossings [47]. This has led to the development of various High Angular Resolution Diffusion Imaging (HARDI) techniques [72] such as Q-Ball Imaging (QBI) or Diffusion Spectrum Imaging (DSI) [73], [74], [76] to explore more precisely the microstructure of biological tissues.

HARDI samples q-space along as many directions as possible in order to reconstruct estimates of the true diffusion probability density function (PDF) – also referred as the Ensemble Average Propagator (EAP) – of water molecules. This true diffusion PDF is model-free and can recover the diffusion of water molecules in any underlying fiber population. HARDI depends on the number of measurements N and the gradient strength (b-value), which will directly affect acquisition time and signal to noise ratio in the signal.

Typically, there are two strategies used in HARDI: 1) sampling of the whole q-space 3D Cartesian grid and estimation of the EAP by inverse Fourier transformation or 2) single shell spherical sampling and estimation of fiber distributions from the diffusion/fiber ODF (QBI), Persistent Angular Structure [58] or Diffusion Orientation Transform [79]. In the first case, a large number of q-space points are taken over the discrete grid ($N > 200$) and the inverse Fourier transform of the measured Diffusion Weighted Imaging (DWI) signal is taken to obtain an estimate of the diffusion PDF. This is Diffusion Spectrum Imaging (DSI) [76], [73], [74]. The method requires very strong imaging gradients ($500 \leq b \leq 20000$ s/mm^2) and a long time for acquisition (15-60 minutes) depending on the number of sampling directions. To infer fiber directions of the diffusion PDF at every voxel, people take an isosurface of the diffusion PDF for a certain radius. Alternatively, they can use the second strategy known as Q-Ball imaging (QBI) i.e just a single shell HARDI acquisition to compute the diffusion orientation distribution function (ODF). With QBI, model-free mathematical approaches can be developed to reconstruct the angular profile of the diffusion displacement probability density function (PDF) of water molecules such as the ODF function which is fundamental in tractography due to the fact that it contains the full angular information of the diffusion PDF and has its maxima aligned with the underlying fiber directions at every voxel.

QBI and the diffusion ODF play a central role in our work related to the development of a robust and linear spherical harmonic estimation of the HARDI signal and to our development of a regularized, fast and robust analytical QBI solution that outperforms the state-of-the-art ODF numerical technique available. Those contributions are fundamental and have already started to impact on the Diffusion MRI, HARDI and Q-Ball Imaging community [50]. They are at the core of our probabilistic and deterministic tractography algorithms devised to best exploit the full distribution of the fiber ODF (see [48], [3] and [49],[4]).

High Order Tensors

Other High Order Tensors (HOT) models to estimate the diffusion function while overcoming the shortcomings of the 2nd order tensor model have also been recently proposed such as the Generalized Diffusion Tensor Imaging (G-DTI) model developed by Ozarslan et al [77], [80] or 4th order Tensor Model [43]. For more details, we refer the reader to our recent article in [53] where we review HOT models and to our article in [7], co-authored with some of our close collaborators, where we review recent mathematical models and computational methods for the processing of Diffusion Magnetic Resonance Images, including state-of-the-art reconstruction of diffusion models, cerebral white matter connectivity analysis, and segmentation techniques.

All these powerful techniques are of utmost importance to acquire a better understanding of the CNS mechanisms and have helped to efficiently tackle and solve a number of important and challenging problems. They have also opened up a landscape of extremely exciting research fields for medicine and neuroscience. Hence, due to the complexity of the CNS data and as the magnetic field strength of scanners increase, as the strength and speed of gradients increase and as new acquisition techniques appear [2], these imaging modalities raise a large amount of mathematical and computational challenges at the core of the research we develop at ATHENA [56].

Improving dMRI Acquisitions and Modeling

One of the most important challenges in diffusion imaging is to improve acquisition schemes and analyse approaches to optimally acquire and accurately represent diffusion profiles in a clinically feasible scanning time. Indeed, a very important and open problem in Diffusion MRI is related to the fact that HARDI scans generally require many times more diffusion gradient than traditional diffusion MRI scan times. This comes at the price of longer scans, which can be problematic for children and people with certain diseases. Patients are usually unable to tolerate long scans and excessive motion of the patient during the acquisition process can force a scan to be aborted or produce useless diffusion MRI images.

Recently, we have developped novel methods for the acquisition and the processing of diffusion magnetic resonance images, to efficiently provide, with just few measurements, new insights into the structure and anatomy of the brain white matter in vivo,

First, we contributed developing real-time reconstruction algorithm based on the Kalman filter [2]. Then, and more recently, we started to explore the utility of Compressive Sensing methods to enable faster acquisition of dMRI data by reducing the number of measurements, while maintaining a high quality for the results. Compressed Sensing (CS) is a recent technique which has been proved to accurately reconstruct sparse signals from undersampled measurements acquired below the Shannon-Nyquist rate. We have also contributed to the reconstruction of the diffusion signal and its important features as the orientation distribution function and the ensemble average propagator, with a special focus on clinical setting in particular for single and multiple Q-shell experiments [10], [11]. Compressive sensing as well as the parametric reconstruction of the diffusion signal in a continuous basis of functions such as the Spherical Polar Fourier basis, have been proved through our recent contributions to be very useful for deriving simple and analytical closed formulae for many important dMRI features, which can be estimated via a reduced number of measurements [10], [11].

We think that such kind of contributions open new perspectives for dMRI applications including, for example, tractography where the improved characterization of the fiber orientations is likely to greatly and quickly help tracking through regions with and/or without crossing fibers [55]

3.2. MEG and EEG

Electroencephalography (EEG) and Magnetoencephalography (MEG) are two non-invasive techniques for measuring (part of) the electrical activity of the brain. While EEG is an old technique (Hans Berger, a German neuropsychiatrist, measured the first human EEG in 1929), MEG is a rather new one: the first measurements of the magnetic field generated by the electrophysiological activity of the brain were made in 1968 at MIT by D. Cohen. Nowadays, EEG is relatively inexpensive and is routinely used to detect and qualify neural activities (epilepsy detection and characterisation, neural disorder qualification, BCI, ...). MEG is, comparatively, much more expensive as SQUIDS only operate under very challenging conditions (at liquid helium temperature) and as a specially shielded room must be used to separate the signal of interest from the ambient noise. However, as it reveals a complementary vision to that of EEG and as it is less sensitive to the head structure, it also bears great hopes and an increasing number of MEG machines are being installed throughout the world. Inria and ODYSSÉE/ATHENA have participated in the acquisition of one such machine installed in the hospital "La Timone" in Marseille.

MEG and EEG can be measured simultaneously (M/EEG) and reveal complementary properties of the electrical fields. The two techniques have temporal resolutions of about the millisecond, which is the typical granularity of the measurable electrical phenomena that arise within the brain. This high temporal resolution makes MEG and EEG attractive for the functional study of the brain. The spatial resolution, on the contrary, is somewhat poor as only a few hundred data points can be acquired simultaneously (about 300-400 for MEG and up to 256 for EEG). MEG and EEG are somewhat complementary with fMRI and SPECT in that those provide a very good spatial resolution but a rather poor temporal resolution (of the order of a second for fMRI and a minute for SPECT). Also, contrarily to fMRI, which "only" measures an haemodynamic response linked to the metabolic demand, MEG and EEG measure a direct consequence of the electrical activity of the brain: it is acknowledged that the signals measured by MEG and EEG correspond to the variations of the post-synaptic

potentials of the pyramidal cells in the cortex. Pyramidal neurons compose approximately 80% of the neurons of the cortex, and it requires at least about 50,000 active such neurons to generate some measurable signal.

While the few hundred temporal curves obtained using M/EEG have a clear clinical interest, they only provide partial information on the localisation of the sources of the activity (as the measurements are made on or outside of the head). Thus the practical use of M/EEG data raises various problems that are at the core of the ATHENA research in this topic:

- First, as acquisition is continuous and is run at a rate up to 1kHz, the amount of data generated by each experiment is huge. Data selection and reduction (finding relevant time blocks or frequency bands) and pre-processing (removing artifacts, enhancing the signal to noise ratio, ...) are largely done manually at present. Making a better and more systematic use of the measurements is an important step to optimally exploit the M/EEG data [1].

- With a proper model of the head and of the sources of brain electromagnetic activity, it is possible to simulate the electrical propagation and reconstruct sources that can explain the measured signal. Proposing better models [6], [9] and means to calibrate them [75] so as to have better reconstructions are other important aims of our work.

- Finally, we wish to exploit the temporal resolution of M/EEG and to apply the various methods we have developed to better understand some aspects of the brain functioning, and/or to extract more subtle information out of the measurements. This is of interest not only as a cognitive goal, but it also serves the purpose of validating our algorithms and can lead to the use of such methods in the field of Brain Computer Interfaces. To be able to conduct such kind of experiments, an EEG lab has been set up at Athena.

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

3. Scientific Foundations

3.1. Mathematical and computational methods

BIOCORE's action is centered on the mathematical modeling of biological systems, more particularly of artificial ecosystems, that have been built or strongly shaped by human. Indeed, the complexity of such systems where the living plays a central role often makes them impossible to understand, control, or optimize without such a formalization. Our theoretical framework of choice for that purpose is Control Theory, whose central concept is "the system", described by state variables, with inputs (action on the system), and outputs (the available measurements on the system). In modeling the ecosystems that we consider, mainly through ordinary differential equations, the state variables are often population, substrate and/or food densities, whose evolution is influenced by the voluntary or involuntary actions of man (inputs and disturbances). The outputs will be some product that one can collect from this ecosystem (harvest, capture, production of a biochemical product, etc), or some measurements (number of individuals, concentrations, etc). Developing a model in biology is however not straightforward: the absence of rigorous laws as in physics, the presence of numerous populations and inputs in the ecosystems, most being irrelevant to the problem at hand, the uncertainties and noise in experiments or even in the biological interactions require the development of techniques to identify and validate the structure of models from data obtained by or with experimentalists.

Building a model is rarely an objective in itself. Once we have checked that it satisfies some biological constraints (eg. densities stay positive) and fitted its parameters to data (requiring tailor-made methods), we perform a mathematical analysis to check that its behavior is consistent with observations. Again, specific methods for this analysis need to be developed that take advantage of the structure of the model (eg. the interactions are monotone) and that take into account the strong uncertainty that is linked to the living, so that qualitative, rather than quantitative, analysis is often the way to go.

In order to act on the system, which often is the purpose of our modeling approach, we then make use of two strong points of Control Theory: 1) the development of observers, that estimate the full internal state of the system from the measurements that we have, and 2) the design of a control law, that imposes to the system the behavior that we want to achieve, be it the regulation at a set point or optimization of its functioning. However, due to the peculiar structure and large uncertainties of our models, we need to develop specific methods. Since actual sensors can be quite costly or simply do not exist, a large part of the internal state often needs to be re-constructed from the measurements and one of the methods we developed consists in integrating the large uncertainties by assuming that some parameters or inputs belong to given intervals. We then developed robust observers that asymptotically estimate intervals for the state variables [7]. Using the directly measured variables and those that have been obtained through such, or other, observers, we then develop control methods that take advantage of the system structure (linked to competition or predation relationships between species in bioreactors or in the trophic networks created or modified by biological control).

3.2. A methodological approach to biology: from genes to ecosystems

One of the objectives of BIOCORE is to develop a methodology that leads to the integration of the different biological levels in our modeling approach: from the biochemical reactions to ecosystems. The regulatory pathways at the cellular level are at the basis of the behavior of the individual organism but, conversely, the external stresses perceived by the individual or population will also influence the intracellular pathways. In a modern "systems biology" view, the dynamics of the whole biosystem/ecosystem emerge from the interconnections among its components, cellular pathways/individual organisms/population. The different scales of size and time that exist at each level will also play an important role in the behavior of the biosystem/ecosystem. The interplay and information transfer between the different levels and scales within a biosystem/ecosystem introduces many new dynamical aspects. We intend to develop methods to understand

the mechanisms at play at each level, from cellular pathways to individual organisms and populations; we assess and model the interconnections and influence between two scale levels (eg., metabolic and genetic; individual organism and population); we explore the possible regulatory and control pathways between two levels; we aim at reducing the size of these large models, in order to isolate subsystems of the main players involved in specific dynamical behaviors.

We develop a theoretical approach of biology by simultaneously considering different levels of description and by linking them, either bottom up (scale transfer) or top down (model reduction). These approaches are used on modeling and analysis of the dynamics of populations of organisms; modeling and analysis of small artificial biological systems using methods of systems biology; control and design of artificial and synthetic biological systems, especially through the coupling of systems.

The goal of this multi level approach is to be able to design or control the cell or individuals to optimize some production or behavior at higher level: for example, control the growth of microalgae via their genetic or metabolic networks, to optimize the production of lipids for bioenergy at the photobioreactor level.

<div align="center">

DEMAR Project-Team

</div>

3. Scientific Foundations

3.1. Modelling and identification of the sensory-motor system

Participants: Mitsuhiro Hayashibe, Christine Azevedo Coste, David Guiraud, Philippe Poignet.

The literature on muscle modelling is vast, but most of research works focus separately on the microscopic and on the macroscopic muscle's functional behaviours. The most widely used microscopic model of muscle contraction was proposed by Huxley in 1957. The Hill-Maxwell macroscopic model was derived from the original model introduced by A.V. Hill in 1938. We may mention the most recent developments including Zahalak's work introducing the distribution moment model that represents a formal mathematical approximation at the sarcomere level of the Huxley cross-bridges model and the works by Bestel and Sorine (2001) who proposed an explanation of the beating of the cardiac muscle by a chemical control input connected to the calcium dynamics in the muscle cells, that stimulates the contractile elements of the model. With respect to this literature, our contributions are mostly linked with the model of the contractile element, through the introduction of the recruitment at the fibre scale formalizing the link between FES parameters, recruitment and Calcium signal path. The resulting controlled model is able to reproduce both short term (twitch) and long term (tetanus) responses. It also matches some of the main properties of the dynamic behaviour of muscles, such as the Hill force-velocity relationship or the instantaneous stiffness of the Mirsky-Parmley model. About integrated functions modelling such as spinal cord reflex loops or central pattern generator, much less groups work on this topic compared to the ones working on brain functions. Mainly neurophysiologists work on this subject and our originality is to combine physiology studies with mathematical modelling and experimental validation using our own neuroprostheses. The same analysis could be drawn with sensory feedback modelling. In this domain, our work is based on the recording and analysis of nerve activity through electro-neurography (ENG). We are interested in interpreting ENG in terms of muscle state in order to feedback useful information for FES controllers and to evaluate the stimulation effect. We believe that this knowledge should help to improve the design and programming of neuroprostheses. We investigate risky but promising fields such as intrafascicular recordings, area on which only few teams in North America (Canada and USA), and Denmark really work on. Very few teams in France, and none at Inria work on the peripheral nervous system modelling, together with experimental protocols that need neuroprostheses. Most of our Inria collaborators work on the central nervous system, except the spinal cord, (ODYSSEE for instance), or other biological functions (SISYPHE for instance). Our contribution concern the following aspects:

- Muscle modelling,
- Sensory organ modelling,
- Electrode nerve interface,
- High level motor function modelling,
- Model parameters identification.

We contribute both to the design of reliable and accurate experiments with a well-controlled environment, to the fitting and implementation of efficient computational methods derived for instance from Sigma Point Kalman Filtering.

3.2. Synthesis and Control of Human Functions

Participants: Christine Azevedo Coste, Philippe Fraisse, Mitsuhiro Hayashibe, David Andreu.

We aim at developping realistic solutions for real clinical problems expressed by patients and medical staff. Different approaches and specifications are developed to answer to those issues in short, mid or long terms. This research axis is therefore obviously strongly related to clinical application objectives. Even though applications can appear very different, the problematic and constraints are usually similar in the context of electrical stimulation: classical desired trajectory tracking is not possible, robustness to disturbances is critical, possible observations of system are limited. Furthermore there is an interaction between body segments under voluntary control of the patient and body segments under artificial control. Finally, this axis relies on modelling and identification results obtained in the first axis and on the technological solutions and approaches developped in the third axis (Neuroprostheses). The robotics framework involved in DEMAR work is close to the tools used and developed by BIPOP team in the context of bipedal robotics. There is no national teams working on those aspects. Within international community, several colleagues carry out researches on the synthesis and control of human functions, most of them belong to the International Functional Electrical Stimulation Society (IFESS) community. In the following we present two sub-objectives. Concerning spinal cord injuries (SCI) context not so many team are now involved in such researches around the world. Our force is to have technological solutions adapted to our theoretical developments. Concerning post-stroke context, several teams in Europe and North America are involved in drop-foot correction using FES. Our team specificity is to have access to the different expertises needed to develop new theoretical and technical solutions: medical expertise, experimental facilities, automatic control expertise, technological developments, industrial partner. These expertises are available in the team and through strong external collaborations.

3.3. Neuroprostheses

Participants: David Andreu, David Guiraud, Guy Cathébras, Fabien Soulier, Serge Bernard.

The main drawbacks of existing implanted FES systems are well known and include insufficient reliability, the complexity of the surgery, limited stimulation selectivity and efficiency, the non-physiological recruitment of motor units and muscle control. In order to develop viable implanted neuroprostheses as palliative solutions for motor control disabilities, the third axis "Neuroprostheses" of our project-team aims at tackling four main challenges: (i) a more physiologically based approach to muscle activation and control, (ii) a fibres' type and localization selective technique and associated technology (iii) a neural prosthesis allowing to make use of automatic control theory and consequently real-time control of stimulation parameters, and (iv) small, reliable, safe and easy-to-implant devices.

Accurate neural stimulation supposes the ability to discriminate fibres' type and localization in nerve and propagation pathway; we thus jointly considered multipolar electrode geometry, complex stimulation profile generation and neuroprosthesis architecture. To face stimulation selectivity issues, the analog output stage of our stimulus generator responds to the following specifications: i) temporal controllability in order to generate current shapes allowing fibres' type and propagation pathway selectivity, ii) spatial controllability of the current applied through multipolar cuff electrodes for fibres' recruitment purposes. We have therefore proposed and patented an original architecture of output current splitter between active poles of a multipolar electrode. The output stage also includes a monotonic DAC (Digital to Analog Converter) by design. However, multipolar electrodes lead to an increasing number of wires between the stimulus generator and the electrode contacts (poles); several research laboratories have proposed complex and selective stimulation strategies involving multipolar electrodes, but they cannot be implanted if we consider multisite stimulation (i.e. stimulating on several nerves to perform a human function as a standing for instance). In contrast, all the solutions tested on humans have been based on centralized implants from which the wires output to only monopolar or bipolar electrodes, since multipolar ones induce to many wires. The only solution is to consider a distributed FES architecture based on communicating controllable implants. Two projects can be cited: Bion technology (main competitor to date), where bipolar stimulation is provided by injectable autonomous units, and the LARSI project, which aimed at multipolar stimulation localized to the sacral roots. In both cases, there was no application breakthrough for reliable standing or walking for paraplegics. The power source, square stimulation shape and bipolar electrode limited the Bion technology, whereas the insufficient selection accuracy of the LARSI implant disqualified it from reliable use.

Keeping the electronics close to the electrode appears to be a good, if not the unique, solution for a complex FES system; this is the concept according to which we direct our neuroprosthesis design and development, in close relationship with other objectives of our project-team (control for instance) but also in close collaboration with medical and industrial partners.

Our efforts are mainly directed to implanted FES system but we also work on surface FES architecture and stimulator; most of our concepts and advancements in implantable neuroprostheses are applicable somehow to external devices.

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

3. Scientific Foundations

3.1. Modelling and simulating microbial ecosystems

Microbial ecosystems naturally put into play phenomena at different scales, from the individual level at a microscopic scale to the population level at a macroscopic scale, with sometimes intermediate levels. The size of substrate molecules is a thousand time smaller than the size of microorganisms and usually diffuse much faster. The substrate consumption of one microorganism is negligible at the population level but the sum of the consumption of its neighbors can modify the local concentration of substrate, which itself modifies microorganism growth, acting as a *feedback loop*. For other variables that change slowly (pH, temperature...) cumulative effects create intermediate time scales, coupling individual and environment dynamics. The very large populations justify macroscopic modelling but for some ecosystems, spatial structures seen at intermediate scale need to be tackled. This is typically the case of biofilm ecosystems, for which the biofilm structure is responsible of characteristics of the overall ecosystem. Models that are purely individual-based or purely populational are rarely truly satisfactory to incorporate current knowledge on microbial ecosystems at various scales and to push ahead mathematical analysis or to derive operational rules.

3.1.1. Macroscopic models

The starting point is the knowledge of biologists that report a large number of mechanisms discovered or shown on laboratory experiments at a population level, such as competition for a growth-limiting substrate, predation interactions, obligate mutualism or communication between bacteria. If each *elementary* mechanism is today well understood and modelled at a macroscopic level, the consideration of several mechanisms together in a single model is still raising several questions of understanding and prediction. This is typically the case when there is more than one growth-limiting substrate in the chemostat model or when one couples species competition with a spatial structure (flocculation, niches...).

1. *Non-spatial models.*
   Ordinary differential equations (ODE) are the common way to describe the evolution of the size or concentration of species populations and their functional contribution in resource transformation (such as substrate degradation) in homogeneous or perfectly mixed compartments (or ecological niches). The well-known chemostat model [87] used in microbiology for single strain:

$$
\begin{aligned}
\dot{s} &= -\tfrac{1}{y}\mu(s)\,b + D\,(s_{in} - s) \\
\dot{b} &= \mu(s)\,b - D\,b
\end{aligned}
$$

   (where s and b stand respectively for the substrate and biomass concentrations, $\mu(s)$ is specific growth function, D the dilution rate, s_{in} the input substrate concentration), has to be extended to cope with the specificity of microbial ecosystems in the following directions.

   – very large number (hundreds or thousands) of species. This leads to the problem of characterization of their distribution during the transients, that is a way to study the *functional redundancy* of ecosystems.

   – environmental fluctuations (input flow rate, input concentration, temperature, pH...). This impacts the efficiency of a microbial ecosystem, when biological and environmental time scales are different. *Singular perturbations* is the technique we use to separate *slow* variables from *fast* ones, leading to approximations of the dynamics on *slow manifolds* to be determined and analyzed.

- interactions due to several limited resources and trophic chains. Most of the literature on the chemostat considers models with single limited resource, while some work studied purely essential or substitutable resources.

- several populations of bacteria (for each species) to describe the effects of certain spatial structures that are artificially created in bioreactors or naturally found in soils, like flocks, colonies or biofilms: the planktonic (or free) cells and the biofilm (or fixed) biomass (for telluric ecosystems, such a distinction is also relevant to represent the sticking/non sticking characteristics of soil). Considering simple models of aggregates (that are not spatialized) can provide a simplified model of the dynamics of the overall biomass.

- *active* and *dormant* bacteria. This distinction is motivated by the observations made on ecosystems of sparse resources such as arid soils.

2. *Spatial models.*
   In the spirit of lattice differential equations, representations in terms of networks of (abstract) interconnected bioreactors propose an intermediate level between models of average biomass (a single ODE) and a continuous representation of space (PDE). A model of interconnected bioreactors is a way to *implicitly* take into account spatial heterogeneity, without requiring a precise knowledge of it. It is similar to the island models used in ecology [85] but coupled with the dynamics of abiotic resources and hydrodynamics laws (transport, percolation, diffusion) governing the transfers between patches. This approach appears to be relevant for telluric ecosystems, for which pedologists report that microbial activities in soil are usually concentrated in *hot-spots* that could be seen as small bioreactors. Understanding the role of the topology of the interconnection network and how a spatial structure impacts the outputs is also relevant in biotechnology to improve the yield or stability of processes.

3.1.2. Microscopic models

In these models (birth and death processes, neutral models, individual-based models) the dynamic of the population is described in terms of discrete events: birth and death of individuals, or jumps in terms of biomass. These models can be gathered under the same framework that could be called *Markov stochastic processes with discrete events*. Most of the time they should be coupled with continuous components like the size of each individual or the dynamic of the resources (represented in terms of ODE or PDE).

The Markovian framework allows on the one hand sharp analyses and rescaling techniques [76]; on the other hand it induces a simplification in the memory structure of the process that is important in terms of simulation. Indeed, as the future state of the system depends from the past only through the present state, only the current state should be kept in memory for simulation.

We will consider three families of processes with discrete events, from simplest to most complex.

1. *Birth and death processes.*
   These models [77] are of first importance in small population size. They indeed allow investigation of near-to-extinction situations in a more realistic ways than the classical ODE models: they permit the computation, analytically but most of the time numerically, the distribution of extinction time and the probability of extinction. Efforts should be made to developed efficient Monte Carlo simulation procedures and approximation techniques for extinction probability and time distribution evaluation. In larger population sizes, they are advantageously approximated by diffusion models (see next section).

2. *The neutral models.*
   In *neutral* models [82] sizes of different species evolve as birth and death processes with immigration: all individuals have the same characteristics and are not spatialized. Such hypotheses could be considered unrealistic from a purely biological perspective, but these models focus on some precise properties to be simulated and predicted (for instance the biodiversity).

Comparing the prediction of species abundance of these models to real observations provides a way to justify or invalidate the neutral hypothesis. Extensions of the neutral model, that was originally introduced for forest ecology, have to be developed in order to better suit the framework of microbial ecology, such as the non constant size of the populations and spatialized variations.

3. *The individual based models.*
   IBM's [68], [69] appear to be well suited to describe colonies or biofilms [88]: in addition to birth, death and movement events, one has to consider *aggregation* and *detachment* events. The mechanisms that lead to the emergence of spatial patterns of colonies, or the formation of biofilms, which adhere to surface via polymers generated by the bacteria under specific hydrodynamics conditions, are not well understood yet. Typically, one can consider that bacteria inside the aggregates are disadvantaged to access the nutrient.

IBM modelling is a convenient way to propose aggregation and detachment mechanisms at the individual level in terms of random events connected to the geometry of the neighborhood, and to compare generated images with microscopic observations (for instance the confocal microscopy).

One has to be aware that few methods are available to study systematically and rigorously the properties of IBM, contrary to models based on differential equations (ODE, PDE...).

3.1.3. Bridges between models

The "theory of a computational model", that combine two kinds of models (typically ODE and IBM) that are different representations of the same objects, relies on two steps: the "program making" and the "theoretical study", in the spirit of the *double modelling* approach (roughly speaking, it consists in grasping the complexity of a IBM by analyzing accurately the consequences of each hypothesis on the macroscopic behavior of the model, building an approximate model of its global dynamics). Two main tools can be considered.

1. *Change of scale.*
   For IBM models (neutral or Markovian), we consider mean field and moments approximation techniques that provide information at the macroscopic (i.e. populational) level, to be compared with macroscopic models. From a birth-and-death process describing the individual level, a renormalisation can provide a stochastic differential equation at a meso-scale. The *diffusion approximation* technique can be understood as a numerical acceleration technique where the number of births and deaths follows a normal law. These stochastic models at meso-scale can provide additional information compared to deterministic models at a macro-scale, such as parameter identifiability or finite time extinction. The price to pay is to give much more conceptual and numerical efforts, that become less relevant for very large populations.

   For PDE models on spatial domains described with regular patterns (such as models of biofilm), the homogenization technique allows to obtain simpler PDE with constant parameters.

2. *The multi-scale modelling.*
   The spatial heterogeneity in microbial ecosystems requires to consider simultaneously several scales:
   – a *physical* scale. In batch processes, nutrient diffusion can be modelled by adapting the heat equation with Dirichlet boundary conditions. In continuous reactors, a convection-diffusion equation with Neumann boundary conditions is considered instead, the speed vector field being provided by the equations of fluid mechanics. The spatial scale used for the discretization is given by diffusion and flow parameters.
   – a *biological* scale, given by the size and mobility of bacteria. Usually, this scale is larger than the physical one (at least in the liquid phase).
   – an *aggregation* scale of colonies or biofilms, even larger, that provides the spatial patterns.

It is always possible to describe all the processes at the smaller common scale and then use a global representation, but this leads to extremely long computation times. The challenge is to manage these overlapping scales together and guarantee the stability of the numerical schemes. This is the goal of the *multi-scale* approaches [84]. For microbial ecosystems, it consists in

1. proposing new representations of the various scales of aggregation of bacteria in a model, taking into account the attachment-detachment processes determined by the local hydro-dynamics conditions. Here, discussions with specialists of fluid mechanics are required.

2. coupling diversity models (e.g. models based on the neutral assumption) with spatial models (that reproduce the patterns observed on images of microscopy) to better understand the link biodiversity/structure.

3. introducing new *control* variables, considered as independent variables, each of them describing a proper scale. For this purpose, we investigate different techniques available to determine such variables: *mean-field* approximation, *singular perturbations*, *unification by limiting layers* or

4. *renormalising*, that aims at detecting invariants among models of different scales.

3.2. Interpreting and analyzing experimental observations

The validation of microbial models on data is rarely a straightforward task, because observations are most of the time not directly related to the variables of the models. Techniques such as abundance spectrum provided by molecular biology or confocal imagery are relatively recent in the field of microbial ecosystems. The signals provided by theses devices leave many research questions open in terms of data interpretation and experiments design. One can distinguish three kinds of key information that are needed at the basis of model assumptions:

- structure of the communities (i.e. who is present?),
- nature of interactions between species (competition, mutualism, syntrophism...),
- spatial structure of the ecosystems.

3.2.1. Assessment of community structures

Ecosystems biodiversity can be observed at different levels, depending on the kind of observations. One usually distinguish:

1. *The taxonomic diversity.* Several techniques developed by molecular biologists can gather information on the genetic structure of communities:
   - *sequencing of a given gene in the community.* The RNA 16S gene is often chosen to identify bacteria or Archeae.
   - *molecular fingerprints.* Some regions in the sequence of the RNA 16S gene encode faithfully the taxa species and can be amplified by PCR techniques.
   - *the sequencing of the overall genetic material of a community* (meta-genomic)

   All these techniques bring new problems of data interpretation to estimate in a robust manner the properties of communities. The signals are combinations of contributions of abundances from each taxon. For an ecosystem with a limited diversity, composed of known species, the signal allows to determine with no ambiguity the abundances. In natural ecosystems, the signal is more complex and it is hopeless to determine uniquely the taxa distribution.

2. *The functional diversity.* It is usually observed at a larger scale, measuring the performances of the overall ecosystem to convert organic matter. The taxonomic diversity does not usually provide such information (it is possible to study *functional genes* but this is much more difficult than studying the 16S one).

A convenient way to study the functional performance of microbial community dynamics is to grow the same microbial community on different substrate compositions, and monitor its performance on these different substrates. Neutral community models [82] provide a reference for what would happen if no functional differences are present in the community. The deviation of experimental observations from neutral model predictions can be considered as a measure of functional diversity.

Understanding the links between taxonomic and functional diversity is currently a tremendous research question in biology about genotype/phenotype links, that one can also find in the specific context of microbial ecosystems.

3.2.2. *Characterization of the interactions*

The role of biodiversity and its preservation in ecosystems are research questions currently largely open in ecology. The nature and number of interactions between bacterial populations are poorly known, and are most probably a key to understand biodiversity. In the classical chemostat model, inter-specific interactions are rarely considered. Also in theoretical ecology, interaction information is typically encoded in an *interaction matrix*, but the coupling with common abiotic resources and the stoichiometry is hardly addressed in the models.

The information provided by confocal microscopy is also a way to estimate the distance of interactions between microorganisms and substrates. This knowledge is not often documented although it is crucial for the construction of IBM.

3.2.3. *Observation of spatial structures*

Schematically, one can distinguish two origins of spatialization:

1. due the physics of the environment. In bioprocesses, this happens typically for large tank size (inducing *dead zones*) or sludge accumulation making the suspension closer from a porous medium than a liquid one. Numerical experimentation can be driven, coupling a solver of the equations of the fluid mechanics with microbiology equations. Then, the spatial distribution of the biomass can be observed and used to calibrate simpler models. Typically, a dead zone is modelled as a diffusive interconnection between two perfect (abstract) tanks.

   But the biotechnology industry aims at considering more sophisticated devices than simple tanks. For instance, the fluidized bed technique consists in creating a counter-current with oxygen bubbles for preventing the biomass to leave the rector. In more complex systems, such as soil ecosystems, it is difficult to obtain faithful simulations because the spatial structure is rarely known with accuracy. Nevertheless, local observations at the level of pores can be achieved, providing information for the construction of models.

2. due to the formation of aggregates (flocks, biofilms...) or biomass wall attachment. Patterns (from ten to a hundred micro-meters) can be observed with confocal microscopy.

   Spatial distribution of bacteria, shape of patterns and composition of the aggregates help to express hypotheses on individual behaviors. But quantification and variability of images provided by confocal microscopy are difficult. An open question is to determine the relevant morphological indicators that characterize aggregation and the formation of biofilms.

3.3. Identifying, controlling and optimizing bioprocesses

The dynamics of the microbial models possess specificities that do not allow the application of the popular methods of the theory of automatic control [71], such as linear control, feedback linearization or canonical forms.

- positivity constraints. State variables, as well as control inputs, have to stay non-negative (input flow pump cannot be reversed because of contamination issues).
- non-linearity. Several models have non-controllable or non-observable linearizations when inhibition effects are present (i.e. change of monotonicity in the growth curves).
- model and measurement uncertainties. In biology, it is rarely relevant to consider model uncertainties as additive Gaussian or finite energy signals.

3.3.1. Software sensors and identification

Sensors in biology are often poor and do not provide the measurements of all the state variables of the models: substrate, strain and product concentrations. In addition, measurements are often spoilt by errors. For instance optical density measurements give an indirect measure of the biomass, influenced by abiotic factors that share the same medium.

Analytical techniques are well suited to ODE models of small dimension, such as:

- guaranteed set-membership observers, when the system is non observable or in presence of unknown inputs,
- (non-linear) changes of coordinates, when the system is observable but not in a canonical form for the construction of observers with exponential convergence.

Software sensors can be also derived with the help of simulation based approaches like particle filtering techniques [75], [74]. This method is suited to diffusion models that approximate birth and death processes. Such softwares will allow us to investigate the different sources of randomness: demography, environment, but mainly imprecision of the sensors.

Similarly, identification techniques for constant parameters are based on sensor models as well as demography and environmental randomness models. In this case, Bayesian and non-Bayesian statistical techniques can be used [73], [70].

3.3.2. Bioprocess stabilization

In bioprocesses, the most efficient bacterial species at steady state are often inhibited by too large concentrations of substrates (this corresponds to assuming that the growth function $s \mapsto \mu(s)$ in the classical chemostat model is non-monotonic). This implies that the washout equilibrium (i.e. disappearance of the biomass) can be attractive, making the bioprocess bi-stable.

A common way to globally stabilize the dynamics toward the efficient equilibrium is to manipulate the dilution rate D [71]. But a diminution of the input flow rate for the stabilization requires to have enough room for an upstream storage, which is an expensive solution especially for developing countries that need to be equipped with new installations.

Alternative ways are proposed to stabilize bioprocesses without restricting the input flow rate:

- either by *physical means*, in terms of recirculation and bypass loops, or membranes as a selective way to keep bacteria and their aggregates inside the tank and improve its efficiency.
- either by *biological means*. The *biological control* consists in adding a small quantity another species with particular growth characteristics, that will help the other species to win the competition in the end.

3.3.3. Optimal control of bioreactors

The filling stage of bioreactors, or "fed-batch", is often time consuming because the quantity of initial biomass is small and consequently the population growth is slow. The minimal time is a typical criterion for designing a filling strategy, but the optimal feedback synthesis is non trivial and may present singular arcs when the growth function is non-monotonic [86].

Recent progress have been made in the consideration of

- multi-species in sequential reactors (having more than one strain makes significantly more difficult to analyze singular arcs because of the higher dimensions of the state space, and there is little literature on the subject),
- energy consumption of flow pumps and the value of byproducts of the biological reactions such as biogas in the criterion (instead of minimal time or as penalties). Recent concerns about sustainable development encourage engineers to look for compromises between those objectives under constraints on output concentrations.

3.3.4. Plant design and optimization

We distinguish two kind of setups:

1. *The industrial setup.* A research question, largely open today, is to identify networks of interconnections of bioreactors that are the most relevant for industrial applications in terms of the following objectives:

   – reasonably simple configurations (i.e. with a limited number of tanks and connections),

   – significant improvement of the residence time at steady state over single or simpler configurations, or shapes of the reservoirs such that the total volume required for a given desired conversion factor at steady state is reduced.

2. *The bioremediation setup.* Typically, the concentration of pollutant in a natural reservoir is solution of a transport-diffusion PDE, but the optimal control of the transport term is almost not studied in the literature.

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

3. Scientific Foundations

3.1. Scientific Foundations

The recent advent of an increasing number of new microscopy techniques giving access to high throughput screenings and micro or nano-metric resolutions provides a means for quantitative imaging of biological structures and phenomena. To conduct quantitative biological studies based on these new data, it is necessary to develop non-standard specific tools. This requires using a multi-disciplinary approach. We need biologists to define experiment protocols and interpret the results, but also physicists to model the sensors, computer scientists to develop algorithms and mathematicians to model the resulting information. These different expertises are combined within the Morpheme team. This generates a fecund frame for exchanging expertise, knowledge, leading to an optimal framework for the different tasks (imaging, image analysis, classification, modeling). We thus aim at providing adapted and robust tools required to describe, explain and model fundamental phenomena underlying the morphogenesis of cellular and supra-cellular biological structures. Combining experimental manipulations, in vivo imaging, image processing and computational modeling, we plan to provide methods for the quantitative analysis of the morphological changes that occur during development. This is of key importance as the morphology and topology of mesoscopic structures govern organ and cell function. Alterations in the genetic programs underlying cellular morphogenesis have been linked to a range of pathologies.

Biological questions we will focus on include:

1. what are the parameters and the factors controlling the establishment of ramified structures? (Are they really organize to ensure maximal coverage? How are genetical and physical constraints limiting their morphology?),

2. how are newly generated cells incorporated into reorganizing tissues during development? (is the relative position of cells governed by the lineage they belong to?)

Our goal is to characterize different populations or development conditions based on the shape of cellular and supra-cellular structures, e.g. micro-vascular networks, dendrite/axon networks, tissues from 2D, 2D+t, 3D or 3D+t images (obtained with confocal microscopy, video-microscopy, photon-microscopy or micro-tomography). We plan to extract shapes or quantitative parameters to characterize the morphometric properties of different samples. On the one hand, we will propose numerical and biological models explaining the temporal evolution of the sample, and on the other hand, we will statistically analyze shapes and complex structures to identify relevant markers for classification purposes. This should contribute to a better understanding of the development of normal tissues but also to a characterization at the supra-cellular scale of different pathologies such as Alzheimer, cancer, diabetes, or the Fragile X Syndrome. In this multidisciplinary context, several challenges have to be faced. The expertise of biologists concerning sample generation, as well as optimization of experimental protocols and imaging conditions, is of course crucial. However, the imaging protocols optimized for a qualitative analysis may be sub-optimal for quantitative biology. Second, sample imaging is only a first step, as we need to extract quantitative information. Achieving quantitative imaging remains an open issue in biology, and requires close interactions between biologists, computer scientists and applied mathematicians. On the one hand, experimental and imag- ing protocols should integrate constraints from the downstream computer-assisted analysis, yielding to a trade-off between qualitative optimized and quantitative optimized protocols. On the other hand, computer analysis should integrate constraints specific to the biological problem, from acquisition to quantitative information extraction. There is therefore a need of specificity for embedding precise biological information for a given task. Besides, a level of generality is also desirable for addressing data from different teams acquired with different protocols and/or sensors. The mathematical modeling of the physics of the acquisition system will yield higher performance reconstruction/restoration algorithms in terms of accuracy. Therefore, physicists and computer scientists have to work together. Quantitative information extraction also has to deal with both the complexity of the structures of interest (e.g., very dense network, small

structure detection in a volume, multiscale behavior, $\cdots$) and the unavoidable defects of in vivo imaging (artifacts, missing data, $\cdots$). Incorporating biological expertise in model-based segmentation methods provides the required specificity while robustness gained from a methodological analysis increases the generality. Finally, beyond image processing, we aim at quantifying and then statistically analyzing shapes and complex structures (e.g., neuronal or vascular networks), static or in evolution, taking into account variability. In this context, learning methods will be developed for determining (dis)similarity measures between two samples or for determining directly a classification rule using discriminative models, generative models, or hybrid models. Besides, some metrics for comparing, classifying and characterizing objects under study are necessary. We will construct such metrics for biological structures such as neuronal or vascular networks. Attention will be paid to computational cost and scalability of the developed algorithms: biological experimentations generally yield huge data sets resulting from high throughput screenings. The research of Morpheme will be developed along the following axes:

- **Imaging:** this includes i) definition of the studied populations (experimental conditions) and preparation of samples, ii) definition of relevant quantitative characteristics and optimized acquisition protocol (staining, imaging, $\cdots$) for the specific biological question, and iii) reconstruction/restoration of native data to improve the image readability and interpretation.

- **Feature extraction:** this consists in detecting and delineating the biological structures of interest from images. Embedding biological properties in the algorithms and models is a key issue. Two main challenges are the variability, both in shape and scale, of biological structures and the huge size of data sets. Following features along time will allow to address morphogenesis and structure development.

- **Classification/Interpretation:** considering a database of images containing different populations, we can infer the parameters associated with a given model on each dataset from which the biological structure under study has been extracted. We plan to define classification schemes for characterizing the different populations based either on the model parameters, or on some specific metric between the extracted structures.

- **Modeling:** two aspects will be considered. This first one consists in modeling biological phenomena such as axon growing or network topology in different contexts. One main advantage of our team is the possibility to use the image information for calibrating and/or validating the biological models. Calibration induces parameter inference as a main challenge. The second aspect consists in using a prior based on biological properties for extracting relevant information from images. Here again, combining biology and computer science expertise is a key point.

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

3. Scientific Foundations

3.1. Neural networks dynamics

The study of neural networks is certainly motivated by the dream to understand how brain is working. But, beyond the comprehension of brain or even of simpler neural systems in less evolved animals, there is also the desire to exhibit general mechanisms or principles at work in the nervous system. One possible strategy is to propose mathematical models of neural activity, at different space and time scales, depending on the type of phenomena under consideration. However, beyond the mere proposal of new models, which can rapidly result in a plethora, there is also a need to understand some fundamental keys ruling the behaviour of neural networks, and, from this, to extract new ideas that can be tested in real experiments. Therefore, there is a need to make a thorough analysis of these models. An efficient approach, developed in our team, consists of analysing neural networks as dynamical systems. This allows to address several issues. A first, natural issue is to ask about the (generic) dynamics exhibited by the system when control parameters vary. This naturally leads to analyse the bifurcations occurring in the network and which phenomenological parameters control these bifurcations. Another issue concerns the interplay between neuron dynamics and synaptic network structure.

In this spirit, our team has been able to characterize the generic dynamics exhibited by models such as conductance-based Integrate and Fire models [2], [64], [65], models of epilepsy [77], effects of synaptic plasticity (see the corresponding section below).
Selected publications on this topic.

3.2. Mean-field approaches

Modelling neural activity at scales integrating the effect of thousands of neurons is of central importance for several reasons. First, most imaging techniques are not able to measure individual neuron activity ("microscopic" scale), but are instead measuring mesoscopic effects resulting from the activity of several hundreds to several hundreds of thousands of neurons. Second, anatomical data recorded in the cortex reveal the existence of structures, such as the cortical columns, with a diameter of about 50μm to 1mm, containing of the order of one hundred to one hundred thousand neurons belonging to a few different species. The description of this collective dynamics requires models which are different from individual neurons models. In particular, when the number of neurons is large enough averaging effects appear, and the collective dynamics is well described by an effective mean-field, summarizing the effect of the interactions of a neuron with the other neurons, and depending on a few effective control parameters. This vision, inherited from statistical physics requires that the space scale be large enough to include a large number of microscopic components (here neurons) and small enough so that the region considered is homogeneous.
Our group is developing mathematical and numerical methods allowing on one hand to produce dynamic mean-field equations from the physiological characteristics of neural structure (neurons type, synapse type and anatomical connectivity between neurons populations), and on the other so simulate those equations. Our investigations have shown that the rigorous dynamics mean-field equations can have a quite more complex structure than the ones commonly used in the literature (e.g. Jansen-Rit, 95) as soon as realistic effects such as synaptic variability are taken into account. Our goal is to relate those theoretical results with experimental measurement, especially in the field of optical imaging. For this we are collaborating with the DYVA team at INT, Marseille.
Selected publications on this topic.

3.3. Neural fields

Neural fields are a phenomenological way of describing the activity of population of neurons by delay integrodifferential equations. This continuous approximation turns out to be very useful to model large brain areas such as those involved in visual perception. The mathematical properties of these equations and their solutions are still imperfectly known, in particular in the presence of different time scales and of noise.
Selected publications on this topic.

3.4. Spike train statistics

The neuronal activity is manifested by the emission of action potentials ("spikes") constituting spike trains. Those spike trains are usually not exactly reproducible when repeating the same experiment, even with a very good control ensuring that experimental conditions have not changed. Therefore, researchers are seeking models for spike train statistics, assumed to be characterized by a hidden probability giving the statistics of spatio-temporal spike patterns. A current goal in experimental analysis of spike trains is to approximate this probability from data. Several approach exist either based on (i) generic principles (maximum likelihood, maximum entropy); (ii) phenomenological models (Linear-Non linear, Generalized Linear Model, mean-field); (iii) Analytical results on spike train statistics in Neural Network models.

Our group is working on those 3 aspects, on a fundamental and on a practical (numerical) level. On one hand, we have published analytical (and rigorous) results on statistics of spike trains in canonical neural network models (Integrate and Fire, conductance based with chemical and electric synapses) [3],[13], [63] . The main result is the characterization of spike train statistics by a Gibbs distribution whose potential can be explicitly computed using some approximations. Note that this result does not require an assumption of stationarity. We have also shown that the distributions considered in the cases (i), (ii), (iii) above are all Gibbs distributions [43], [12]. On the other hand we are proposing new algorithms for data processing [22]. We have developed a C++ software for spike train statistics based on Gibbs distributions analysis and freely available at http://enas.gforge.inria.fr/v3/. We are using this software in collaboration with several biologist groups involved in the analysis of retina spike trains (Centro de Neurociencia Valparaiso; Molecular Biology Lab, Princeton; Institut de la vision, Paris) [26], [22], [55].
Selected publications on this topic.

3.5. Synaptic Plasticity

Neural networks show amazing abilities for information storage and processing, and stimulus-dependent activity shaping, to evolve and adapt. These capabilities are mainly conditioned by plasticity mechanisms, and especially synaptic plasticity, inducing a mutual coupling between network structure and neuron dynamics. Synaptic plasticity occurs at many levels of organization and time scales in the nervous system (Bienenstock, Cooper, and Munroe, 1982). It is of course involved in memory and learning mechanisms, but it also alters excitability of brain areas and regulates behavioural states (e.g. transition between sleep and wakeful activity). Therefore, understanding the effects of synaptic plasticity on neurons dynamics is a crucial challenge. On experimental grounds, different synaptic plasticity mechanisms have been exhibited from the Hebbian's ones (Hebb, 1949) to Long Term Potentiation (LTP) and Long Term Depression (LTD), and more recently to Spike Time Dependent Plasticity (STDP) (Markram, Lubke, Frotscher, and Sakmann, 1997; Bi and Poo, 2001). Synaptic plasticity implies that activity guides the way synapses evolve; but the resulting connectivity structure in turn can raise new dynamical regimes. This interaction becomes even more complex if the considered basic architecture is not feed-forward but includes recurrent synaptic links, like in cortical structures. Understanding this mutual coupling between dynamics and topology and its effects on the computations made by the network is a key problem in computational neuroscience.

Our group is developing mathematical and numerical methods to analyse this mutual interaction. Especially, we have shown that plasticity mechanisms, Hebbian-like or STDP, have strong effects on neuron dynamics complexity, such as dynamics complexity reduction, and spike statistics (convergence to a specific Gibbs distribution via a variational principle), resulting in a response-adaptation of the network to learned stimuli [73], [74],[4]. Also, we are currently studying the conjugated effects of synaptic and intrinsic plasticity in collaboration with H. Berry (Inria Beagle) and B. Delord, J. Naudé, ISIR team, Paris.
Selected publications on this topic.

3.6. Visual neuroscience

Our group focuses on the visual system to understand how information is encoded and processed resulting in visual percepts. To do so, we propose functional models of the visual system using a variety of mathematical formalisms, depending on the scale at which models are built, such as spiking neural networks or neural fields. So far, our efforts have been mainly focused on the study of retinal processing and motion integration at the level of V1 and MT cortical areas.

At the retina level, we are modelling its circuitry [6] and we are studying the statistics of the spike train output (see, e.g., the software ENAS http://enas.gforge.inria.fr/v3/). Real cell recordings are also analysed in collaboration with Institut de la Vision (Paris). At the level of V1-MT cortical areas, we are investigating the temporal dynamics of motion integration for a wide range of visual stimuli [38], [41], [23], [39], [42] [75], [62][5]. This work is done in collaboration with Institut de Neuroscience de la Timone (Marseille).
Selected publications on this topic.

3.7. Neuromorphic vision

From the simplest vision architectures in insects to the extremely complex cortical hierarchy in primates, it is fascinating to observe how biology has found efficient solutions to solve vision problems. Pioneers in computer vision had this dream to build machines that could match and perhaps outperform human vision. This goal has not been reached, at least not on the scale that was originally planned, but the field of computer vision has met many other challenges from an unexpected variety of applications and fostered entirely new scientific and technological areas such as computer graphics and medical image analysis. However, modelling and emulating with computers biological vision largely remains an open challenge while there are still many outstanding issues in computer vision.

Our group is working on neuromorphic vision by proposing bio-inspired methods following our progress in visual neuroscience. Our goal is to bridge the gap between biological and computer vision, by applying our visual neuroscience models to challenging problems from computer vision such as optical flow estimation [76], coding/decoding approaches [20], [21], [37] [69], [68] or classification [14] [66].
Selected publications on this topic.

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

3. Scientific Foundations

3.1. Analysis of structures resulting from meristem activity

To analyze plant growth and structure, we focus mainly on methods for analyzing sequences and tree-structured data. Theses methods range from algorithms for computing distance between sequences or tree-structured data to statistical models.

- *Combinatorial approaches*: plant structures exhibit complex branching organizations of their organs like internodes, leaves, shoots, axes, branches, etc. These structures can be analyzed with combinatorial methods in order to compare them or to reveal particular types of organization. We investigate a family of techniques to quantify distances between branching systems based on non-linear structural alignment (similar to edit-operation methods used for sequence comparison). Based on these techniques, we study the notion of (topology-based) self-similarity of branching structures in order to define a notion of degree of redundancy for any tree structure and to quantify in this way botanical notions, such as the physiological states of a meristem, fundamental to the description of plant morphogenesis.

- *Statistical modeling*: We investigate different categories of statistical models corresponding to different types of structures.

  - Longitudinal data corresponding to plant growth follow up: the statistical models of interest are equilibrium renewal processes and generalized linear mixed models for longitudinal count data.

  - Repeated patterns within sequences or trees: the statistical models of interest are mainly (hidden) variable-order Markov chains. Hidden variable-order Markov chains were in particular applied to characterize permutation patterns in phyllotaxis and the alternation between flowering and vegetative growth units along sympodial tree axes.

  - Homogeneous zones (or change points) within sequences or trees: most of the statistical models of interest are hidden Markovian models (hidden semi-Markov chains, semi-Markov switching linear mixed models and semi-Markov switching generalized linear models for sequences and different families of hidden Markov tree models). A complementary approach consists in applying multiple change-point models. The branching structure of a parent shoot is often organized as a succession of branching zones while the succession of shoot at the more macroscopic scale exhibit roughly stationary phases separated by marked change points.

  We investigate both estimation methods and diagnostic tools for these different categories of models. In particular we focus on diagnostic tools for latent structure models (e.g. hidden Markovian models or multiple change-point models) that consist in exploring the latent structure space.

- *A new generation of morphogenesis models*: Designing morphogenesis models of the plant development at the macroscopic scales is a challenging problem. As opposed to modeling approaches that attempt to describe plant development on the basis of the integration of purely mechanistic models of various plant functions, we intend to design models that tightly couple mechanistic and empirical sub-models that are elaborated in our plant architecture analysis approach. Empirical models are used as a powerful complementary source of knowledge in places where knowledge about mechanistic processes is lacking or weak. We chose to implement such integrated models in a programming language dedicated to dynamical systems with dynamical structure $(DS)^2$, such as L-systems or MGS. This type of language plays the role of an integration framework for sub-models of heterogeneous nature.

3.2. Meristem functioning and development

In this second scientific axis, we develop models of meristem growth at tissue level in order to integrate various sources of knowledge and to analyze their dynamic and complex spatial interaction. To carry out this integration, we need to develop a complete methodological approach containing:

- algorithms for the automatized segmentation in 3D, and cell lineage tracking throughout time, for images coming from confocal microscopy,

- design of high-level routines and user interfaces to distribute these image analysis tools to the scientific community,

- tools for structural and statistical analysis of 3D meristem structure (spatial statistics, multiscale geometric and topological analysis),

- physical models of cells interactions based on spring-mass systems or on tensorial mechanics at the level of cells,

- models of biochemical networks of hormonal and gene driven regulation, at the cellular and tissue level, using continuous and discrete formalisms,

- and models of cell development taking into account the effects of growth and cell divisions on the two previous classes of models.

3.3. OpenAlea: An open-software platform for plant modeling

OpenAlea is a component based, open-software platform for interdisciplinary research in plant modeling and simulation. This platform is used for the integration and comparison of different models and tools provided by the research community. It is based on the Python (http://www.python.org) language that aims at being both a *glue* language for the different modules and an efficient modeling language for developing new models and tools. *OpenAlea* currently includes modules for plant simulation, analysis and modeling at different scales (*V-Plants* modules), for modeling ecophysiological processes such as radiative transfer, transpiration and photosynthesis (*RATP*, *Caribu*, *Adel*, *TopVine*, *Ecomeristem*) and for 3D visualization of plant architecture at different scales (*PlantGL*).

OpenAlea is the result of a collaborative effort associating 20 french research teams in plant modeling from Inria, CIRAD, INRA, LaBRI, Laboratory Jean Kuntzmann and ENS Lyon. The Virtual Plants team coordinates both development and modeling consortiums, and is more particularly in charge of the development of the kernel and of some of the main data structures such as multi-scale tree graphs and statistical sequences.

OpenAlea is a fundamental tool to share models and methods in interdisciplinary research (comprising botany, ecophysiology, forestry, agronomy, applied mathematics and computer science approaches). Embedded in Python and its scientific libraries, the platform may be used as a flexible and useful toolbox by biologists and modelers for various purposes (research, teaching, rapid model prototyping, communication, etc.).

FOCUS Project-Team

3. Scientific Foundations

3.1. Models

The objective of Focus is to develop concepts, techniques, and possibly also tools, that may contribute to the analysis and synthesis of CBUS. Fundamental to these activities is *modeling*. Therefore designing, developing and studying computational models appropriate for CBUS is a central activity of the project. The models are used to formalize and verify important computational properties of the systems, as well as to propose new linguistic constructs.

The models we study are in the process calculi (e.g., the π-calculus) and λ-calculus tradition. Such models, with their emphasis on algebra, well address compositionality—a central property in our approach to problems. Accordingly, the techniques we employ are mainly operational techniques based on notions of behavioral equivalence, and techniques based on algebra, mathematical logics, and type theory.

The sections below provide some more details on why process calculi, λ-calculi, and related techniques, should be useful for CBUS.

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

3. Scientific Foundations

3.1. Parallelism, concurrency, and distribution

Concurrency management is at the heart of diffuse programming. Since the execution platforms are highly heterogeneous, many different concurrency principles and models may be involved. Asynchronous concurrency is the basis of shared-memory process handling within multiprocessor or multicore computers, of direct or fifo-based message passing in distributed networks, and of fifo- or interrupt-based event handling in web-based human-machine interaction or sensor handling. Synchronous or quasi-synchronous concurrency is the basis of signal processing, of real-time control, and of safety-critical information acquisition and display. Interfacing existing devices based on these different concurrency principles within HOP or other diffuse programming languages will require better understanding of the underlying concurrency models and of the way they can nicely cooperate, a currently ill-resolved problem.

3.2. Web and functional programming

We are studying new paradigms for programming Web applications that rely on multi-tier functional programming [4]. We have created a Web programming environment named HOP. It relies on a single formalism for programming the server-side and the client-side of the applications as well as for configuring the execution engine.

HOP is a functional language based on the SCHEME programming language. That is, it is a strict functional language, fully polymorphic, supporting side effects, and dynamically type-checked. HOP is implemented as an extension of the BIGLOO compiler that we develop [5]. In the past, we have extensively studied static analyses (type systems and inference, abstract interpretations, as well as classical compiler optimizations) to improve the efficiency of compilation in both space and time.

3.3. Security of diffuse programs

The main goal of our security research is to provide scalable and rigorous language-based techniques that can be integrated into multi-tier compilers to enforce the security of diffuse programs. Research on language-based security has been carried on before in former Inria teams [2], [1]. In particular previous research has focused on controlling information flow to ensure confidentiality.

Typical language-based solutions to these problems are founded on static analysis, logics, provable cryptography, and compilers that generate correct code by construction [3]. Relying on the multi-tier programming language HOP that tames the complexity of writing and analysing secure diffuse applications, we are studying language-based solutions to prominent web security problems such as code injection and cross-site scripting, to name a few.

LOGNET Team (section vide)

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

3. Scientific Foundations

3.1. Scientific Foundations

The main mathematical tools and formalisms used in MAESTRO include:

- theory of stochastic processes: Markov process, renewal process, point process, Palm measure, large deviations, branching process, mean-field approximation;

- theory of dynamical discrete-event systems: queues, fluid approximation;

- theory of control and scheduling: dynamic programming, Markov decision process, game theory, deterministic and stochastic scheduling, pathwise comparison;

- theory of singular perturbations;

- random matrix theory.

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

3. Scientific Foundations

3.1. Scientific Foundations

The project develops tools and theory in the following domains: Discrete Mathematics (in particular Graph Theory), Algorithmics, Combinatorial Optimization and Simulation.

Typically, a telecommunication network (or an interconnection network) is modeled by a graph. A vertex may represent either a processor or a router or any of the following: a switch, a radio device, a site or a person. An edge (or arc) corresponds to a connection between the elements represented by the vertices (logical or physical connection). We can associate more information both to the vertices (for example what kind of switch is used, optical or not, number of ports, equipment cost) and to the edges (weights which might correspond to length, cost, bandwidth, capacity) or colors (modeling either wavelengths or frequencies or failures) etc. Depending on the application, various models can be defined and have to be specified. This modeling part is an important task. To solve the problems, we manage, when possible, to find polynomial algorithms. For example, a maximum set of disjoint paths between two given vertices is by Menger's theorem equal to the minimum cardinality of a cut. This problem can be solved in polynomial time using graph theoretic tools or flow theory or linear programming. On the contrary, determining whether in a directed graph there exists a pair of disjoint paths, one from s_1 to t_1 and the other from s_2 to t_2, is an NP-complete problem, and so are all the problems which aim at minimizing the cost of a network which can satisfy certain traffic requirements. In addition to deterministic hypotheses (for example if a connection fails it is considered as definitely down and not intermittently), the project started recently to consider probabilistic ones.

Graph coloring is an example of concept which appears in various contexts: WDM networks where colors represent wavelengths, radio networks where colors represent frequencies, fault tolerance where colors represent shared risk resource groups, and scheduling problems. Another tool concerns the development of new algorithmic aspects like parameterized algorithms.

OASIS Project-Team

3. Scientific Foundations

3.1. Programming with distributed objects and components

The paradigm of object-oriented programming, although not very recent, is clearly still not properly defined and implemented; for example notions like inheritance, sub-typing or overloading have as many definitions as there are different object languages. The introduction of concurrency and distribution into objects also increases the complexity. It appeared that standard Java constituents such as RMI (Remote Method Invocation) do not help building, in a transparent way, sequential, multi-threaded, or distributed applications. Indeed allowing, as RMI does, the execution of the same application to proceed on a shared-memory multiprocessors architecture as well as on a network of computing units (intranet, Internet), or on any hierarchical combination of both, is not sufficient for providing a convenient and reliable programming environment.

The question is thus: how to ease the construction (i.e. programming), deployment and evolution of distributed applications ?

One of the answers we suggest relies on the concept of active object, that acts as a single entity, abstraction of a thread, a set of objects and a location. Active objects communicate by asynchronous method calls thanks to the use of futures. ProActive is a Java library that implements this notion of active objects. ProActive can also be seen as a middleware supporting deployment, runtime support, and efficient communication for large scale distributed applications.

Another answer we provide relies on component-oriented programming. In particular, we have defined parallel and hierarchical distributed components starting from the Fractal component model developed by Inria and France-Telecom [58]. We have been involved in the design of the Grid Component Model (GCM) [4], which is one of the major results produced by the CoreGrid European Network of Excellence. The GCM has been standardized at ETSI ([64] for the last published standard), and most of our research on component models are related to it. On the practical side, ProActive/GCM is the implementation of the GCM above the ProActive programming library.

We have developed over time skills in both theoretical and applicative side fields, such as distribution, fault-tolerance, verification, etc., to provide a better programming and runtime environment for object oriented and component oriented applications.

3.2. Formal models for distributed objects

A few years ago, we designed the ASP calculus [7] for modelling distributed objects. It remains to this date one of our major scientific foundations. ASP is a calculus for distributed objects interacting using asynchronous method calls with generalized futures. Those futures naturally come with a transparent and automatic synchronisation called wait-by-necessity. In large-scale systems, our approach provides both a good structure and a strong decoupling between threads, and thus scalability. Our work on ASP provides very generic results on expressiveness and determinism, and the potential of this approach has been further demonstrated by its capacity to cope with advanced issues, such as mobility, group communications, and components [6].

ASP provides confluence and determinism properties for distributed objects. Such results should allow one to program parallel and distributed applications that behave in a deterministic manner, even if they are distributed over local or wide area networks.

The ASP calculus is a model for the ProActive library. An extension of ASP has been built to model distributed asynchronous components. A functional fragment of ASP has been modelled in the Isabelle theorem prover [8].

3.3. Verification, static analysis, and model-checking

Even with the help of high-level libraries, distributed systems are more difficult to program than classical applications. The complexity of interactions and synchronisations between remote parts of a system increases the difficulty of analysing their behaviours. Consequently, safety, security, or liveness properties are particularly difficult to ensure for these applications. Formal verification of software systems has been active for a long time, but its impact on the development methodology and tools has been slower than in the domain of hardware and circuits. This is true both at a theoretical and at a practical level; our contributions include:

- the definition of adequate models representing programs,
- the mastering of state complexity through abstraction techniques, new algorithmic approaches, or research on advanced parallel or distributed verification methods,
- the design of software tools that hide to the final user the complexity of the underlying theory.

We concentrate on the area of distributed component systems, where we get better descriptions of the structure of the system, making the analysis more tractable, but we also find out new interesting problems. For instance, we contributed to a better analysis of the interplay between the functional definition of a component and its possible runtime transformations, expressed by the various management controllers of the component system.

Our approach is bi-directional: from models to program, or back. We use techniques of static analysis and abstract interpretation to extract models from the code of distributed applications, or from dedicated specification formalisms [3]. On the other hand, we generate "safe by construction" code skeletons, from high level specifications; this guarantees the behavioural properties of the components. We then use generic tools from the verification community to check properties of these models. We concentrate on behavioural properties, expressed in terms of temporal logics (safety, liveness), of adequacy of an implementation to its specification and of correct composition of software components.

PLANETE Project-Team

3. Scientific Foundations

3.1. Experimental approach to Networking

Based on a practical view, the Planète approach to address the above research topics is to design new communication protocols or mechanisms, to implement and to evaluate them either by simulation or by experimentation on real network platforms (such as PlanetLab and OneLab). Our work includes a substantial technological component since we implement our mechanisms in pre-operational systems and we also develop applications that integrate the designed mechanisms as experimentation and demonstration tools. We also work on the design and development of networking experimentation tools such as the ns-3 network simulator and experimental platforms. We work in close collaboration with research and development industrial teams.

In addition to our experimentation and deployment specificities, we closely work with researchers from various domains to broaden the range of techniques we can apply to networks. In particular, we apply techniques of the information and queuing theories to evaluate the performance of protocols and systems. The collaboration with physicists and mathematicians is, from our point of view, a promising approach to find solutions that will build the future of the Internet.

In order to carry out our approach as well as possible, it is important to attend and contribute to IETF (Internet Engineering Task Force) and other standardization bodies meetings on a regular basis, in order to propose and discuss our ideas in the working groups related to our topics of interests.

AXIS Project-Team (section vide)

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

3. Scientific Foundations

3.1. Probabilistic approaches

Following a Bayesian methodology as far as possible, probabilistic models are used within the AYIN team for three purposes: to describe the class of images to be expected from any given scene, to describe prior knowledge about the scene and to incorporate specific constraints. The models used in AYIN fall into the following two classes.

3.1.1. *Markov random fields*

Markov random fields were introduced to image processing in the Eighties, and were quickly applied to the full range of inverse problems in computer vision. They owe their popularity to their flexible and intuitive nature, which makes them an ideal modelling tool, and to the existence of standard and easy-to-implement algorithms for their solution. In the AYIN team, attention is focused on their use in image modelling, in particular of textures; on the development of improved prior models for segmentation; and on the lightening of the heavy computational load traditionally associated with these techniques, in particular via the study of varieties of hierarchical random fields.

3.1.2. *Stochastic geometry*

One of the grand challenges of computer vision and image processing is the expression and use of prior geometric information. For satellite and aerial imagery, this problem has become increasingly important as the increasing resolution of the data results in the necessity to model geometric structures hitherto invisible. One of the most promising approaches to the inclusion of this type of information is stochastic geometry, which is an important line of research in the AYIN team. Instead of defining probabilities for different types of image, probabilities are defined for configurations of an indeterminate number of interacting, parameterized objects located in the image. Such probability distribution are called 'marked point processes'. Such processes have been recently applied to skin care problems.

3.2. Parameter estimation

One of the most important problems studied in the AYIN team is how to estimate the parameters that appear in the models. For probabilistic models, the problem is quite easily framed, but is not necessarily easy to solve, particularly in the case when it is necessary to extract simultaneously both the information of interest and the parameters from the data.

3.3. Hierarchical models

Another line of research in the AYIN team concerns development of graph-based, in particular, hierarchical models for very high resolution image analysis and classification. A specific hierarchical model recently developed in AYIN represents an image by a forest structure, where leaf nodes represent image regions at the finest level of partition, while other nodes correspond to image regions at the coarser levels of partitions. The AYIN team is interested in developing multi-feature models of image regions as an ensemble of spectral, texture, geometrical and classification features, and establishing new criteria for comparing image regions. Recent research concerns extension of hierarchical models to a temporal dimension, for analyzing multitemporal data series.

<div align="center">

COPRIN Project-Team

</div>

3. Scientific Foundations

3.1. Interval analysis

We are interested in real-valued system solving ($f(X) = 0$, $f(X) \leq 0$), in optimization problems, and in the proof of the existence of properties (for example, it exists X such that $f(X) = 0$ or it exist two values X_1, X_2 such that $f(X_1) > 0$ and $f(X_2) < 0$). There are few restrictions on the function f as we are able to manage explicit functions using classical mathematical operators (e.g. $\sin(x + y) + \log(\cos(e^x) + y^2)$ as well as implicit functions (e.g. determining if there are parameter values of a parametrized matrix such that the determinant of the matrix is negative, without calculating the analytical form of the determinant).

Solutions are searched within a finite domain (called a *box*) which may be either continuous or mixed (i.e. for which some variables must belong to a continuous range while other variables may only have values within a discrete set). An important point is that we aim at finding all the solutions within the domain whenever the computer arithmetic will allow it: in other words we are looking for *certified* solutions. For example, for 0-dimensional system solving, we will provide a box that contains one, and only one, solution together with a numerical approximation of this solution. This solution may further be refined at will using multi-precision.

The core of our methods is the use of *interval analysis* that allows one to manipulate mathematical expressions whose unknowns have interval values. A basic component of interval analysis is the *interval evaluation* of an expression. Given an analytical expression F in the unknowns $\{x_1, x_2, ..., x_n\}$ and ranges $\{X_1, X_2, ..., X_n\}$ for these unknowns we are able to compute a range $[A, B]$, called the interval evaluation, such that

$$\forall \{x_1, x_2, ..., x_n\} \in \{X_1, X_2, ..., X_n\}, A \leq F(x_1, x_2, ..., x_n) \leq B \tag{5}$$

In other words the interval evaluation provides a lower bound of the minimum of F and an upper bound of its maximum over the box.

For example if $F = x \, sin(x + x^2)$ and $x \in [0.5, 1.6]$, then $F([0.5, 1.6]) = [-1.362037441, 1.6]$, meaning that for any x in [0.5,0.6] we guarantee that $-1.362037441 \leq f(x) \leq 1.6$.

The interval evaluation of an expression has interesting properties:

- it can be implemented in such a way that the results are guaranteed with respect to round-off errors i.e. property 1 is still valid in spite of numerical errors induced by the use of floating point numbers
- if $A > 0$ or $B < 0$, then no values of the unknowns in their respective ranges can cancel F
- if $A > 0$ ($B < 0$), then F is positive (negative) for any value of the unknowns in their respective ranges

A major drawback of the interval evaluation is that $A(B)$ may be overestimated i.e. values of $x_1, x_2, ..., x_n$ such that $F(x_1, x_2, ..., x_n) = A(B)$ may not exist. This overestimation occurs because in our calculation each occurrence of a variable is considered as an independent variable. Hence if a variable has multiple occurrences, then an overestimation may occur. Such phenomena can be observed in the previous example where $B = 1.6$ while the real maximum of F is approximately 0.9144. The value of B is obtained because we are using in our calculation the formula $F = xsin(y + z^2)$ with y, z having the same interval value than x.

Fortunately there are methods that allow one to reduce the overestimation and the overestimation amount decreases with the width of the ranges. The latter remark leads to the use of a branch-and-bound strategy in which for a given box a variable range will be bisected, thereby creating two new boxes that are stored in a list and processed later on. The algorithm is complete if all boxes in the list have been processed, or if during the process a box generates an answer to the problem at hand (e.g. if we want to prove that $F(X) < 0$, then the algorithm stops as soon as $F(\mathcal{B}) \geq 0$ for a certain box $\mathcal{B}$).

A generic interval analysis algorithm involves the following steps on the current box [1], [7], [5]:

1. *exclusion operators*: these operators determine that there is no solution to the problem within a given box. An important issue here is the extensive and smart use of the monotonicity of the functions
2. *filters*: these operators may reduce the size of the box i.e. decrease the width of the allowed ranges for the variables [11], [19]
3. *existence operators*: they allow one to determine the existence of a unique solution within a given box and are usually associated with a numerical scheme that allows for the computation of this solution in a safe way
4. *bisection*: choose one of the variable and bisect its range for creating two new boxes
5. *storage*: store the new boxes in the list

The scope of the COPRIN project is to address all these steps in order to find the most efficient procedures. Our efforts focus on mathematical developments (adapting classical theorems to interval analysis, proving interval analysis theorems), the use of symbolic computation and formal proofs (a symbolic pre-processing allows one to automatically adapt the solver to the structure of the problem), software implementation and experimental tests (for validation purposes).

3.2. Robotics

COPRIN has a long-standing tradition of robotics studies, especially for closed-loop robots [4]. We address theoretical issues with the purpose of obtaining analytical and theoretical solutions, but in many cases only numerical solutions can be obtained due to the complexity of the problem. This approach has motivated the use of interval analysis for two reasons:

1. the versatility of interval analysis allows us to address issues (e.g. singularity analysis) that cannot be tackled by any other method due to the size of the problem
2. uncertainties (which are inherent to a robotic device) have to be taken into account so that the *real* robot is guaranteed to have the same properties as the *theoretical* one, even in the worst case. This is a crucial issue for many applications in robotics (e.g. medical or assistance robot)

Our field of study in robotics focuses on *kinematic* issues [13], [21] such as workspace and singularity analysis, positioning accuracy [24], trajectory planning, reliability, calibration [33], modularity management and, prominently, *appropriate design*, i.e. determining the dimensioning of a robot mechanical architecture that guarantees that the real robot satisfies a given set of requirements [28]. The methods that we develop can be used for other robotic problems, see for example the management of uncertainties in aircraft design [34], [10].

Our theoretical work must be validated through experiments that are essential for the sake of credibility. A contrario, experiments will feed theoretical work. Hence COPRIN works with partners on the development of real robots but also develops its own prototypes. We usually develop a new robot prototype every 6 years but since 2008 we have started the development of seven new robot prototypes, mostly related to assistance robotics. Furthermore we have extended our development to devices that are not strictly robots but are part of an overall environment for assistance. We benefit here from the development of new miniature, low energy computers with an interface for analog and logical sensors such as the Arduino or the Phidgets. We intend to make a full use of such devices, especially for assistance purpose

In term of applications we have focused up to now on the development of special machines (machine-tool, ultra-high accuracy positioning device, spatial telescope). Although this activity will be pursued, we have started in 2008 a long-term move toward *service robotics*, i.e. robots that are closer to human activity. In service robotics we are interested in domotics, smart objects, rehabilitation and medical robots [8], [9], [26] and entertainment, that can be regrouped under the name of *assistance robotics* (see section 6.2.1.3). Compared to special machines for which pricing is not an issue (up to a certain point), cost is an important element for assistance robotics. While we plan to develop simple robotic systems using only standard hardware, our work will focus on a different issue: *adaptability*. We aim at providing assistance devices that are adapted to the end-user, its trajectory of life and its environment, are easy to install (because installation uncertainties are taken into account at the design stage), have a low intrusivity and are guaranteed to fulfill a set of requirements.

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

3. Scientific Foundations

3.1. Logic-based Knowledge Representation and Reasoning

We follow the mainstream *logical* approach to the KRR domain. First-order logic (FOL) is the reference logic in KRR and most formalisms in this area can be translated into fragments (i.e., particular subsets) of FOL. A large part of research in this domain can be seen as studying the *trade-off* between the expressivity of languages and the complexity of (sound and complete) reasoning in these languages. The fundamental problem in KRR languages is entailment checking: is a given piece of knowledge entailed by other pieces of knowledge, for instance from a knowledge base (KB)? Another important problem is *consistency* checking: is a set of knowledge pieces (for instance the knowledge base itself) consistent, i.e., is it sure that nothing absurd can be entailed from it? The *query answering* problem is a topical problem (see Sect. 3.3). It asks for the set of answers to a query in the KB. In the special case of boolean queries (i.e., queries with a yes/no answer), it can be recast as entailment checking.

3.2. Graph-based Knowledge Representation and Reasoning

Besides logical foundations, we are interested in KRR formalisms that comply, or aim at complying with the following requirements: to have good *computational* properties and to allow users of knowledge-based systems to have a maximal *understanding and control* over each step of the knowledge base building process and use.

These two requirements are the core motivations for our specific approach to KRR, which is based on labelled *graphs*. Indeed, we view labelled graphs as an *abstract representation* of knowledge that can be expressed in many KRR languages (different kinds of conceptual graphs —historically our main focus—, the Semantic Web language RDF (Resource Description Framework), its extension RDFS (RDF Schema) expressive rules equivalent to the so-called tuple-generating-dependencies in databases, some description logics dedicated to query answering, etc.). For these languages, reasoning can be based on the structure of objects, thus based on graph-theoretic notions, while staying logically founded.

More precisely, our basic objects are labelled graphs (or hypergraphs) representing entities and relationships between these entities. These graphs have a natural translation in first-order logic. Our basic reasoning tool is graph homomorphism. The fundamental property is that graph homomorphism is sound and complete with respect to logical entailment i.e. given two (labelled) graphs G and H, there is a homomorphism from G to H *if and only if* the formula assigned to G is entailed by the formula assigned to H. In other words, logical reasonings on these graphs can be performed by graph mechanisms. These knowledge constructs and the associated reasoning mechanisms can be extended (to represent rules for instance) while keeping this fundamental correspondence between graphs and logics.

3.3. Ontological Query Answering

Querying knowledge bases is a central problem in knowledge representation and in database theory. A knowledge base (KB) is classically composed of a terminological part (metadata, ontology) and an assertional part (facts, data). Queries are supposed to be at least as expressive as the basic queries in databases, i.e., conjunctive queries, which can be seen as existentially closed conjunctions of atoms or as labelled graphs. The challenge is to define good trade-offs between the expressivity of the ontological language and the complexity of querying data in presence of ontological knowledge. Classical ontogical languages, typically description logics, were not designed for efficient querying. On the other hand, database languages were able to process complex queries on huge databases, but without taking the ontology into account. There is thus a need for new languages and mechanisms, able to cope with the ever growing size of knowledge bases in the Semantic Web or in scientific domains.

This problem is related to two other problems identified as fundamental in KRR:

- *Query-answering with incomplete information.* Incomplete information means that it might be unknown whether a given assertion is true or false. Databases classically make the so-called closed-world assumption: every fact that cannot be retrieved or inferred from the base is assumed to be false. Knowledge bases classically make the open-world assumption: if something cannot be inferred from the base, and neither can its negation, then its truth status is unknown. The need of coping with incomplete information is a distinctive feature of querying knowledge bases with respect to querying classical databases (however, as explained above, this distinction tends to disappear). The presence of incomplete information makes the query answering task much more difficult.

- *Reasoning with rules.* Researching types of rules and adequate manners to process them is a mainstream topic in the Semantic Web, and, more generally a crucial issue for knowledge-based systems. For several years, we have been studying some rules, both in their logical and their graph form, which are syntactically very simple but also very expressive. These rules can be seen as an abstraction of ontological knowledge expressed in the main languages used in the context of KB querying. See Sect. 6.1 for details on the results obtained.

A problem generalising the above described problems, and particularly relevant in the context of multiple data/metadata sources, is *querying hybrid knowledge bases*. In a hybrid knowledge base, each component may have its own formalism and its own reasoning mechanisms. There may be a common ontology shared by all components, or each component may have its own ontology, with mappings being defined among the ontologies. The question is what kind of interactions between these components and/or what limitations on the languages preserve the decidability of basic problems and if so, a "reasonable"complexity. Note that there are strong connections with data integration in databases.

3.4. Imperfect Information and Priorities

While classical FOL is the kernel of many KRR languages, to solve real-world problems we often need to consider features that cannot be expressed purely (or not naturally) in classical logic. The logic- and graph-based formalisms used for previous points have thus to be extended with such features. The following requirements have been identified from scenarios in decision making in the agronomy domain (see Sect. 4.2):

1. to cope with vague and uncertain information and preferences in queries;
2. to cope with multi-granularity knowledge;
3. to take into account different and potentially conflicting viewpoints ;
4. to integrate decision notions (priorities, gravity, risk, benefit);
5. to integrate argumentation-based reasoning.

Although the solutions we will develop need to be validated on the applications that motivated them, we also want them to be sufficiently generic to be applied in other contexts. One angle of attack (but not the only possible one) consists in increasing the expressivity of our core languages, while trying to preserve their essential combinatorial properties, so that algorithmic optimizations can be transferred to these extensions. To achieve that goal, our main research directions are: non-monotonic reasonings (see ANR project ASPIQ in Sect. 8.1), as well as argumentation and preferences (see Sect. 6.2).

LAGADIC Project-Team

3. Scientific Foundations

3.1. Visual servoing

Basically, visual servoing techniques consist in using the data provided by one or several cameras in order to control the motions of a dynamic system [1]. Such systems are usually robot arms, or mobile robots, but can also be virtual robots, or even a virtual camera. A large variety of positioning tasks, or mobile target tracking, can be implemented by controlling from one to all the degrees of freedom of the system. Whatever the sensor configuration, which can vary from one on-board camera on the robot end-effector to several free-standing cameras, a set of visual features has to be selected at best from the image measurements available, allowing to control the desired degrees of freedom. A control law has also to be designed so that these visual features $\mathbf{s}(t)$ reach a desired value $\mathbf{s}^*$, defining a correct realization of the task. A desired planned trajectory $\mathbf{s}^*(t)$ can also be tracked. The control principle is thus to regulate to zero the error vector $\mathbf{s}(t) - \mathbf{s}^*(t)$. With a vision sensor providing 2D measurements, potential visual features are numerous, since 2D data (coordinates of feature points in the image, moments, ...) as well as 3D data provided by a localization algorithm exploiting the extracted 2D features can be considered. It is also possible to combine 2D and 3D visual features to take the advantages of each approach while avoiding their respective drawbacks.

More precisely, a set $\mathbf{s}$ of k visual features can be taken into account in a visual servoing scheme if it can be written:

$$\mathbf{s} = \mathbf{s}(\mathbf{x}(\mathbf{p}(t)), \mathbf{a}) \tag{6}$$

where $\mathbf{p}(t)$ describes the pose at the instant t between the camera frame and the target frame, $\mathbf{x}$ the image measurements, and $\mathbf{a}$ a set of parameters encoding a potential additional knowledge, if available (such as for instance a coarse approximation of the camera calibration parameters, or the 3D model of the target in some cases).

The time variation of $\mathbf{s}$ can be linked to the relative instantaneous velocity $\mathbf{v}$ between the camera and the scene:

$$\dot{\mathbf{s}} = \frac{\partial \mathbf{s}}{\partial \mathbf{p}} \, \dot{\mathbf{p}} = \mathbf{L_s} \, \mathbf{v} \tag{7}$$

where $\mathbf{L_s}$ is the interaction matrix related to $\mathbf{s}$. This interaction matrix plays an essential role. Indeed, if we consider for instance an eye-in-hand system and the camera velocity as input of the robot controller, we obtain when the control law is designed to try to obtain an exponential decoupled decrease of the error:

$$\mathbf{v}_c = -\lambda \widehat{\mathbf{L_s}}^{+} (\mathbf{s} - \mathbf{s}^*) - \widehat{\mathbf{L_s}}^{+} \widehat{\frac{\partial \mathbf{s}}{\partial t}} \tag{8}$$

where λ is a proportional gain that has to be tuned to minimize the time-to-convergence, $\widehat{\mathbf{L_s}}^{+}$ is the pseudo-inverse of a model or an approximation of the interaction matrix, and $\widehat{\frac{\partial \mathbf{s}}{\partial t}}$ an estimation of the features velocity due to a possible own object motion.

From the selected visual features and the corresponding interaction matrix, the behavior of the system will have particular properties as for stability, robustness with respect to noise or to calibration errors, robot 3D trajectory, etc. Usually, the interaction matrix is composed of highly non linear terms and does not present any decoupling properties. This is generally the case when s is directly chosen as x. In some cases, it may lead to inadequate robot trajectories or even motions impossible to realize, local minimum, tasks singularities, etc. It is thus extremely important to design adequate visual features for each robot task or application, the ideal case (very difficult to obtain) being when the corresponding interaction matrix is constant, leading to a simple linear control system. To conclude in few words, **visual servoing is basically a non linear control problem. Our Holy Grail quest is to transform it into a linear control problem.**

Furthermore, embedding visual servoing in the task function approach allows solving efficiently the redundancy problems that appear when the visual task does not constrain all the degrees of freedom of the system. It is then possible to realize simultaneously the visual task and secondary tasks such as visual inspection, or joint limits or singularities avoidance. This formalism can also be used for tasks sequencing purposes in order to deal with high level complex applications.

3.2. Visual tracking

Elaboration of object tracking algorithms in image sequences is an important issue for researches and applications related to visual servoing and more generally for robot vision. A robust extraction and real time spatio-temporal tracking process of visual cues is indeed one of the keys to success of a visual servoing task. If fiducial markers may still be useful to validate theoretical aspects in modeling and control, natural scenes with non cooperative objects and subject to various illumination conditions have to be considered for addressing large scale realistic applications.

Most of the available tracking methods can be divided into two main classes: feature-based and model-based. The former approach focuses on tracking 2D features such as geometrical primitives (points, segments, circles,...), object contours, regions of interest...The latter explicitly uses a model of the tracked objects. This can be either a 3D model or a 2D template of the object. This second class of methods usually provides a more robust solution. Indeed, the main advantage of the model-based methods is that the knowledge about the scene allows improving tracking robustness and performance, by being able to predict hidden movements of the object, detect partial occlusions and acts to reduce the effects of outliers. The challenge is to build algorithms that are fast and robust enough to meet our applications requirements. Therefore, even if we still consider 2D features tracking in some cases, our researches mainly focus on real-time 3D model-based tracking, since these approaches are very accurate, robust, and well adapted to any class of visual servoing schemes. Furthermore, they also meet the requirements of other classes of application, such as augmented reality.

3.3. Slam

Most of the applications involving mobile robotic systems (ground vehicles, aerial robots, automated submarines,...) require a reliable localization of the robot in its environment. A challenging problem is when neither the robot localization nor the map is known. Localization and mapping must then be considered concurrently. This problem is known as Simultaneous Localization And Mapping (Slam). In this case, the robot moves from an unknown location in an unknown environment and proceeds to incrementally build up a navigation map of the environment, while simultaneously using this map to update its estimated position.

Nevertheless, solving the Slam problem is not sufficient for guaranteeing an autonomous and safe navigation. The choice of the representation of the map is, of course, essential. The representation has to support the different levels of the navigation process: motion planning, motion execution and collision avoidance and, at the global level, the definition of an optimal strategy of displacement. The original formulation of the Slam problem is purely metric (since it basically consists in estimating the Cartesian situations of the robot and a set of landmarks), and it does not involve complex representations of the environment. However, it is now well recognized that **several complementary representations are needed to perform exploration, navigation, mapping, and control tasks successfully. We propose to use composite models of the environment that**

mix topological, metric, and grid-based representations. Each type of representation is well adapted to a particular aspect of autonomous navigation: the metric model allows one to locate the robot precisely and plan Cartesian paths, the topological model captures the accessibility of different sites in the environment and allows a coarse localization, and finally the grid representation is useful to characterize the free space and design potential functions used for reactive obstacle avoidance. However, ensuring the consistency of these various representations during the robot exploration, and merging observations acquired from different viewpoints by several cooperative robots, are difficult problems. This is particularly true when different sensing modalities are involved. New studies to derive efficient algorithms for manipulating the hybrid representations (merging, updating, filtering...) while preserving their consistency are needed.

REVES Project-Team

3. Scientific Foundations

3.1. Rendering

We consider plausible rendering to be a first promising research direction, both for images and for sound. Recent developments, such as point rendering, image-based modeling and rendering, and work on the simulation of aging indicate high potential for the development of techniques which render *plausible* rather than extremely accurate images. In particular, such approaches can result in more efficient renderings of very complex scenes (such as outdoors environments). This is true both for visual (image) and sound rendering. In the case of images, such techniques are naturally related to image- or point-based methods. It is important to note that these models are becoming more and more important in the context of network or heterogeneous rendering, where the traditional polygon-based approach is rapidly reaching its limits. Another research direction of interest is realistic rendering using simulation methods, both for images and sound. In some cases, research in these domains has reached a certain level of maturity, for example in the case of lighting and global illumination. For some of these domains, we investigate the possibility of technology transfer with appropriate partners. Nonetheless, certain aspects of these research domains, such as visibility or high-quality sound still have numerous and interesting remaining research challenges.

3.1.1. Plausible Rendering

3.1.1.1. Alternative representations for complex geometry

The key elements required to obtain visually rich simulations, are sufficient geometric detail, textures and lighting effects. A variety of algorithms exist to achieve these goals, for example displacement mapping, that is the displacement of a surface by a function or a series of functions, which are often generated stochastically. With such methods, it is possible to generate convincing representations of terrains or mountains, or of non-smooth objects such as rocks. Traditional approaches used to represent such objects require a very large number of polygons, resulting in slow rendering rates. Much more efficient rendering can be achieved by using point or image based rendering, where the number of elements used for display is view- or image resolution-dependent, resulting in a significant decrease in geometric complexity. Such approaches have very high potential. For example, if all object can be rendered by points, it could be possible to achieve much higher quality local illumination or shading, using more sophisticated and expensive algorithms, since geometric complexity will be reduced. Such novel techniques could lead to a complete replacement of polygon-based rendering for complex scenes. A number of significant technical challenges remain to achieve such a goal, including sampling techniques which adapt well to shading and shadowing algorithms, the development of algorithms and data structures which are both fast and compact, and which can allow interactive or real-time rendering. The type of rendering platforms used, varying from the high-performance graphics workstation all the way to the PDA or mobile phone, is an additional consideration in the development of these structures and algorithms. Such approaches are clearly a suitable choice for network rendering, for games or the modelling of certain natural object or phenomena (such as vegetation, e.g. Figure 1 , or clouds). Other representations merit further research, such as image or video based rendering algorithms, or structures/algorithms such as the "render cache" [37], which we have developed in the past, or even volumetric methods. We will take into account considerations related to heterogeneous rendering platforms, network rendering, and the appropriate choices depending on bandwith or application. Point- or image-based representations can also lead to novel solutions for capturing and representing real objects. By combining real images, sampling techniques and borrowing techniques from other domains (e.g., computer vision, volumetric imaging, tomography etc.) we hope to develop representations of complex natural objects which will allow rapid rendering. Such approaches are closely related to texture synthesis and image-based modeling. We believe that such methods will not replace 3D (laser or range-finger) scans, but could be complementary, and represent a simpler and lower cost alternative for certain applications (architecture, archeology etc.). We are also investigating methods for

adding "natural appearance" to synthetic objects. Such approaches include *weathering* or *aging* techniques, based on physical simulations [27], but also simpler methods such as accessibility maps [34]. The approaches we intend to investigate will attempt to both combine and simplify existing techniques, or develop novel approaches founded on generative models based on observation of the real world.

3.1.1.2. Plausible audio rendering

Similar to image rendering, plausible approaches can be designed for audio rendering. For instance, the complexity of rendering high order reflections of sound waves makes current geometrical approaches inappropriate. However, such high order reflections drive our auditory perception of "reverberation" in a virtual environment and are thus a key aspect of a plausible audio rendering approach. In complex environments, such as cities, with a high geometrical complexity, hundreds or thousands of pedestrians and vehicles, the acoustic field is extremely rich. Here again, current geometrical approaches cannot be used due to the overwhelming number of sound sources to process. We study approaches for statistical modeling of sound scenes to efficiently deal with such complex environments. We also study perceptual approaches to audio rendering which can result in high efficiency rendering algorithms while preserving visual-auditory consistency if required.

Figure 1. Plausible rendering of an outdoors scene containing points, lines and polygons [26], representing a scene with trees, grass and flowers. We can achieve 7-8 frames per second compared to tens of seconds per image using standard polygonal rendering.

3.1.2. High Quality Rendering Using Simulation

3.1.2.1. Non-diffuse lighting

A large body of global illumination research has concentrated on finite element methods for the simulation of the diffuse component and stochastic methods for the non-diffuse component. Mesh-based finite element approaches have a number of limitations, in terms of finding appropriate meshing strategies and form-factor calculations. Error analysis methodologies for finite element and stochastic methods have been very different in the past, and a unified approach would clearly be interesting. Efficient rendering, which is a major advantage of finite element approaches, remains an overall goal for all general global illumination research. For certain cases, stochastic methods can be efficient for all types of light transfers, in particular if we require a view-dependent solution. We are also interested both in *pure* stochastic methods, which do not use finite element techniques. Interesting future directions include filtering for improvement of final image quality as well as beam tracing type approaches [35] which have been recently developed for sound research.

3.1.2.2. Visibility and Shadows

Visibility calculations are central to all global illumination simulations, as well as for all rendering algorithms of images and sound. We have investigated various global visibility structures, and developed robust solutions for scenes typically used in computer graphics. Such analytical data structures [31], [30], [29] typically have robustness or memory consumption problems which make them difficult to apply to scenes of realistic size. Our solutions to date are based on general and flexible formalisms which describe all visibility event in terms of generators (vertices and edges); this approach has been published in the past [28]. Lazy evaluation, as well as hierarchical solutions, are clearly interesting avenues of research, although are probably quite application dependent.

3.1.2.3. Radiosity

For purely diffuse scenes, the radiosity algorithm remains one of the most well-adapted solutions. This area has reached a certain level of maturity, and many of the remaining problems are more technology-transfer oriented We are interested in interactive or real-time renderings of global illumination simulations for very complex scenes, the "cleanup" of input data, the use of application-dependent semantic information and mixed representations and their management. Hierarchical radiosity can also be applied to sound, and the ideas used in clustering methods for lighting can be applied to sound.

3.1.2.4. High-quality audio rendering

Our research on high quality audio rendering is focused on developing efficient algorithms for simulations of geometrical acoustics. It is necessary to develop techniques that can deal with complex scenes, introducing efficient algorithms and data structures (for instance, beam-trees [32] [35]), especially to model early reflections or diffractions from the objects in the environment. Validation of the algorithms is also a key aspect that is necessary in order to determine important acoustical phenomena, mandatory in order to obtain a high-quality result. Recent work by Nicolas Tsingos at Bell Labs [33] has shown that geometrical approaches can lead to high quality modeling of sound reflection and diffraction in a virtual environment (Figure 2). We will pursue this research further, for instance by dealing with more complex geometry (e.g., concert hall, entire building floors).
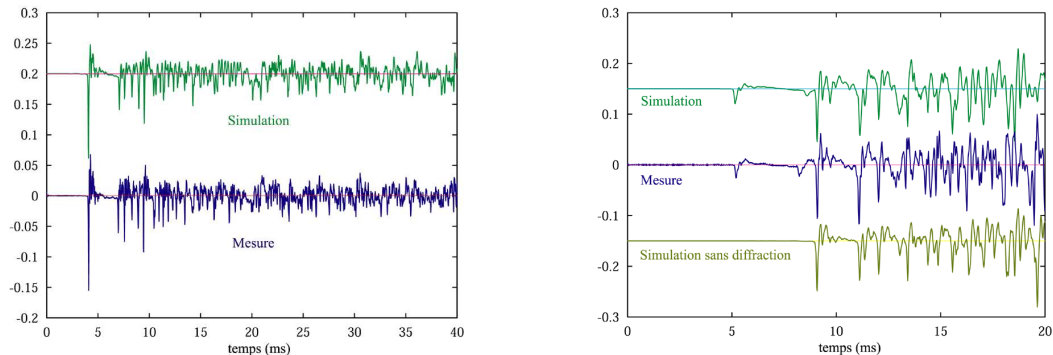

Figure 2. A comparison between a measurement (left) of the sound pressure in a given location of the "Bell Labs Box", a simple test environment built at Bell Laboratories, and a high-quality simulation based on a beam-tracing engine (right). Simulations include effects of reflections off the walls and diffraction off a panel introduced in the room.

Finally, several signal processing issues remain in order to properly and efficiently restitute a 3D soundfield to the ears of the listener over a variety of systems (headphones, speakers). We would like to develop an open and general-purpose API for audio rendering applications. We already completed a preliminary version of a software library: AURELI [36].

STARS Team

3. Scientific Foundations

3.1. Introduction

Stars follows three main research directions: perception for activity recognition, semantic activity recognition, and software engineering for activity recognition. **These three research directions are interleaved**: the software architecture direction provides new methodologies for building safe activity recognition systems and the perception and the semantic activity recognition directions provide new activity recognition techniques which are designed and validated for concrete video analytics and healthcare applications. Conversely, these concrete systems raise new software issues that enrich the software engineering research direction.

Transversally, we consider a new research axis in machine learning, combining a priori knowledge and learning techniques, to set up the various models of an activity recognition system. A major objective is to automate model building or model enrichment at the perception level and at the understanding level.

3.2. Perception for Activity Recognition

Participants: Guillaume Charpiat, François Brémond, Sabine Moisan, Monique Thonnat.

Computer Vision; Cognitive Systems; Learning; Activity Recognition.

3.2.1. *Introduction*

Our main goal in perception is to develop vision algorithms able to address the large variety of conditions characterizing real world scenes in terms of sensor conditions, hardware requirements, lighting conditions, physical objects, and application objectives. We have also several issues related to perception which combine machine learning and perception techniques: learning people appearance, parameters for system control and shape statistics.

3.2.2. *Appearance models and people tracking*

An important issue is to detect in real-time physical objects from perceptual features and predefined 3D models. It requires finding a good balance between efficient methods and precise spatio-temporal models. Many improvements and analysis need to be performed in order to tackle the large range of people detection scenarios.

Appearance models. In particular, we study the temporal variation of the features characterizing the appearance of a human. This task could be achieved by clustering potential candidates depending on their position and their reliability. This task can provide any people tracking algorithms with reliable features allowing for instance to (1) better track people or their body parts during occlusion, or to (2) model people appearance for re-identification purposes in mono and multi-camera networks, which is still an open issue. The underlying challenge of the person re-identification problem arises from significant differences in illumination, pose and camera parameters. The re-identification approaches have two aspects: (1) establishing correspondences between body parts and (2) generating signatures that are invariant to different color responses. As we have already several descriptors which are color invariant, we now focus more on aligning two people detections and on finding their corresponding body parts. Having detected body parts, the approach can handle pose variations. Further, different body parts might have different influence on finding the correct match among a whole gallery dataset. Thus, the re-identification approaches have to search for matching strategies. As the results of the re-identification are always given as the ranking list, re-identification focuses on learning to rank. "Learning to rank" is a type of machine learning problem, in which the goal is to automatically construct a ranking model from a training data.

Therefore, we work on information fusion to handle perceptual features coming from various sensors (several cameras covering a large scale area or heterogeneous sensors capturing more or less precise and rich information). New 3D sensors (e.g. Kinect) are also investigated, to help in getting an accurate segmentation for specific scene conditions.

Long term tracking. For activity recognition we need robust and coherent object tracking over long periods of time (often several hours in videosurveillance and several days in healthcare). To guarantee the long term coherence of tracked objects, spatio-temporal reasoning is required. Modelling and managing the uncertainty of these processes is also an open issue. In Stars we propose to add a reasoning layer to a classical Bayesian framework1pt modelling the uncertainty of the tracked objects. This reasoning layer can take into account the a priori knowledge of the scene for outlier elimination and long-term coherency checking.

Controling system parameters. Another research direction is to manage a library of video processing programs. We are building a perception library by selecting robust algorithms for feature extraction, by insuring they work efficiently with real time constraints and by formalizing their conditions of use within a program supervision model. In the case of video cameras, at least two problems are still open: robust image segmentation and meaningful feature extraction. For these issues, we are developing new learning techniques.

3.2.3. *Learning shape and motion*

Another approach, to improve jointly segmentation and tracking, is to consider videos as 3D volumetric data and to search for trajectories of points that are statistically coherent both spatially and temporally. This point of view enables new kinds of statistical segmentation criteria and ways to learn them.

We are also using the shape statistics developed in [5] for the segmentation of images or videos with shape prior, by learning local segmentation criteria that are suitable for parts of shapes. This unifies patch-based detection methods and active-contour-based segmentation methods in a single framework. These shape statistics can be used also for a fine classification of postures and gestures, in order to extract more precise information from videos for further activity recognition. In particular, the notion of shape dynamics has to be studied.

More generally, to improve segmentation quality and speed, different optimization tools such as graph-cuts can be used, extended or improved.

3.3. Semantic Activity Recognition

Participants: Guillaume Charpiat, François Brémond, Sabine Moisan, Monique Thonnat.

Activity Recognition, Scene Understanding,Computer Vision

3.3.1. *Introduction*

Semantic activity recognition is a complex process where information is abstracted through four levels: signal (e.g. pixel, sound), perceptual features, physical objects and activities. The signal and the feature levels are characterized by strong noise, ambiguous, corrupted and missing data. The whole process of scene understanding consists in analysing this information to bring forth pertinent insight of the scene and its dynamics while handling the low level noise. Moreover, to obtain a semantic abstraction, building activity models is a crucial point. A still open issue consists in determining whether these models should be given a priori or learned. Another challenge consists in organizing this knowledge in order to capitalize experience, share it with others and update it along with experimentation. To face this challenge, tools in knowledge engineering such as machine learning or ontology are needed.

Thus we work along the two following research axes: high level understanding (to recognize the activities of physical objects based on high level activity models) and learning (how to learn the models needed for activity recognition).

3.3.2. *High Level Understanding*

A challenging research axis is to recognize subjective activities of physical objects (i.e. human beings, animals, vehicles) based on a priori models and objective perceptual measures (e.g. robust and coherent object tracks).

To reach this goal, we have defined original activity recognition algorithms and activity models. Activity recognition algorithms include the computation of spatio-temporal relationships between physical objects. All the possible relationships may correspond to activities of interest and all have to be explored in an efficient way. The variety of these activities, generally called video events, is huge and depends on their spatial and temporal granularity, on the number of physical objects involved in the events, and on the event complexity (number of components constituting the event).

Concerning the modelling of activities, we are working towards two directions: the uncertainty management for representing probability distributions and knowledge acquisition facilities based on ontological engineering techniques. For the first direction, we are investigating classical statistical techniques and logical approaches. We have also built a language for video event modelling and a visual concept ontology (including color, texture and spatial concepts) to be extended with temporal concepts (motion, trajectories, events ...) and other perceptual concepts (physiological sensor concepts ...).

3.3.3. *Learning for Activity Recognition*

Given the difficulty of building an activity recognition system with a priori knowledge for a new application, we study how machine learning techniques can automate building or completing models at the perception level and at the understanding level.

At the understanding level, we are learning primitive event detectors. This can be done for example by learning visual concept detectors using SVMs (Support Vector Machines) with perceptual feature samples. An open question is how far can we go in weakly supervised learning for each type of perceptual concept (i.e. leveraging the human annotation task). A second direction is to learn typical composite event models for frequent activities using trajectory clustering or data mining techniques. We name composite event a particular combination of several primitive events.

3.3.4. *Activity Recognition and Discrete Event Systems*

The previous research axes are unavoidable to cope with the semantic interpretations. However they tend to let aside the pure event driven aspects of scenario recognition. These aspects have been studied for a long time at a theoretical level and led to methods and tools that may bring extra value to activity recognition, the most important being the possibility of formal analysis, verification and validation.

We have thus started to specify a formal model to define, analyze, simulate, and prove scenarios. This model deals with both absolute time (to be realistic and efficient in the analysis phase) and logical time (to benefit from well-known mathematical models providing re-usability, easy extension, and verification). Our purpose is to offer a generic tool to express and recognize activities associated with a concrete language to specify activities in the form of a set of scenarios with temporal constraints. The theoretical foundations and the tools being shared with Software Engineering aspects, they will be detailed in section 3.4 .

The results of the research performed in perception and semantic activity recognition (first and second research directions) produce new techniques for scene understanding and contribute to specify the needs for new software architectures (third research direction).

3.4. Software Engineering for Activity Recognition

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, François Brémond.

Software Engineering, Generic Components, Knowledge-based Systems, Software Component Platform, Object-oriented Frameworks, Software Reuse, Model-driven Engineering

The aim of this research axis is to build general solutions and tools to develop systems dedicated to activity recognition. For this, we rely on state-of-the art Software Engineering practices to ensure both sound design and easy use, providing genericity, modularity, adaptability, reusability, extensibility, dependability, and maintainability.

This research requires theoretical studies combined with validation based on concrete experiments conducted in Stars. We work on the following three research axes: models (adapted to the activity recognition domain), platform architecture (to cope with deployment constraints and run time adaptation), and system verification (to generate dependable systems). For all these tasks we follow state of the art Software Engineering practices and, if needed, we attempt to set up new ones.

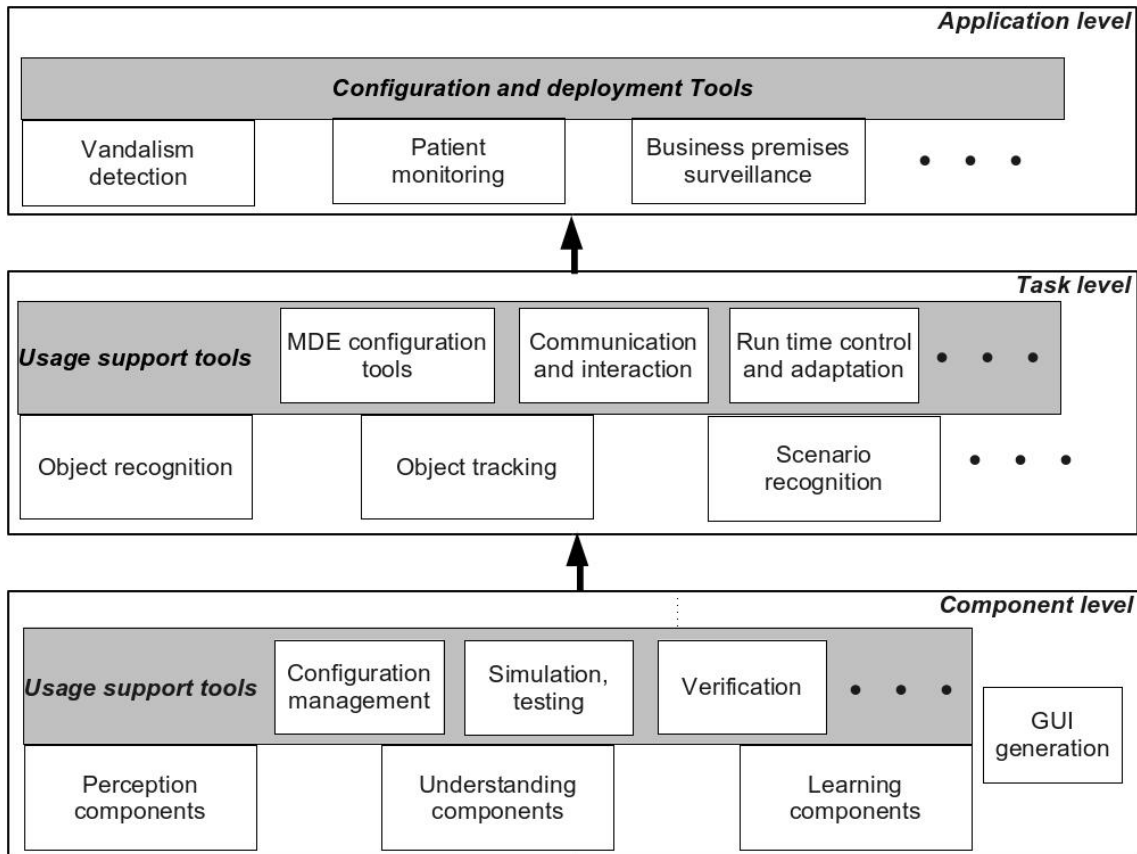### 3.4.1. Platform Architecture for Activity Recognition

Figure 4. Global Architecture of an Activity Recognition The grey areas contain software engineering support modules whereas the other modules correspond to software components (at Task and Component levels) or to generated systems (at Application level).

In the former project teams Orion and Pulsar, we have developed two platforms, one (VSIP), a library of real-time video understanding modules and another one, LAMA [15], a software platform enabling to design not only knowledge bases, but also inference engines, and additional tools. LAMA offers toolkits to build and to adapt all the software elements that compose a knowledge-based system or a cognitive system.

Figure 4 presents our conceptual vision for the architecture of an activity recognition platform. It consists of three levels:

- The **Component Level**, the lowest one, offers software components providing elementary operations and data for perception, understanding, and learning.

- Perception components contain algorithms for sensor management, image and signal analysis, image and video processing (segmentation, tracking...), etc.

- Understanding components provide the building blocks for Knowledge-based Systems: knowledge representation and management, elements for controlling inference engine strategies, etc.

- Learning components implement different learning strategies, such as Support Vector Machines (SVM), Case-based Learning (CBL), clustering, etc.

An Activity Recognition system is likely to pick components from these three packages. Hence, tools must be provided to configure (select, assemble), simulate, verify the resulting component combination. Other support tools may help to generate task or application dedicated languages or graphic interfaces.

- The **Task Level**, the middle one, contains executable realizations of individual tasks that will collaborate in a particular final application. Of course, the code of these tasks is built on top of the components from the previous level. We have already identified several of these important tasks: Object Recognition, Tracking, Scenario Recognition... In the future, other tasks will probably enrich this level.

  For these tasks to nicely collaborate, communication and interaction facilities are needed. We shall also add MDE-enhanced tools for configuration and run-time adaptation.

- The **Application Level** integrates several of these tasks to build a system for a particular type of application, e.g., vandalism detection, patient monitoring, aircraft loading/unloading surveillance, etc.. Each system is parametrized to adapt to its local environment (number, type, location of sensors, scene geometry, visual parameters, number of objects of interest...). Thus configuration and deployment facilities are required.

The philosophy of this architecture is to offer at each level a balance between the widest possible genericity and the maximum effective reusability, in particular at the code level.

To cope with real application requirements, we shall also investigate distributed architecture, real time implementation, and user interfaces.

Concerning implementation issues, we shall use when possible existing open standard tools such as NuSMV for model-checking, Eclipse for graphic interfaces or model engineering support, Alloy for constraint representation and SAT solving, etc. Note that, in Figure 4 , some of the boxes can be naturally adapted from SUP existing elements (many perception and understanding components, program supervision, scenario recognition...) whereas others are to be developed, completely or partially (learning components, most support and configuration tools).

3.4.2. Discrete Event Models of Activities

As mentioned in the previous section (3.3) we have started to specify a formal model of scenario dealing with both absolute time and logical time. Our scenario and time models as well as the platform verification tools rely on a formal basis, namely the synchronous paradigm. To recognize scenarios, we consider activity descriptions as synchronous reactive systems and we apply general modelling methods to express scenario behaviour.

Activity recognition systems usually exhibit many safeness issues. From the software engineering point of view we only consider software security. Our previous work on verification and validation has to be pursued; in particular, we need to test its scalability and to develop associated tools. Model-checking is an appealing technique since it can be automatized and helps to produce a code that has been formally proved. Our verification method follows a compositional approach, a well-known way to cope with scalability problems in model-checking.

Moreover, recognizing real scenarios is not a purely deterministic process. Sensor performance, precision of image analysis, scenario descriptions may induce various kinds of uncertainty. While taking into account this uncertainty, we should still keep our model of time deterministic, modular, and formally verifiable. To formally describe probabilistic timed systems, the most popular approach involves probabilistic extension of timed automata. New model checking techniques can be used as verification means, but relying on model checking techniques is not sufficient. Model checking is a powerful tool to prove decidable properties but introducing uncertainty may lead to infinite state or even undecidable properties. Thus model checking validation has to be completed with non exhaustive methods such as abstract interpretation.

3.4.3. *Model-Driven Engineering for Configuration and Control and Control of Video Surveillance systems*

Model-driven engineering techniques can support the configuration and dynamic adaptation of video surveillance systems designed with our SUP activity recognition platform. The challenge is to cope with the many—functional as well as nonfunctional—causes of variability both in the video application specification and in the concrete SUP implementation. We have used *feature models* to define two models: a generic model of video surveillance applications and a model of configuration for SUP components and chains. Both of them express variability factors. Ultimately, we wish to automatically generate a SUP component assembly from an application specification, using models to represent transformations [54]. Our models are enriched with intra- and inter-models constraints. Inter-models constraints specify models to represent transformations. Feature models are appropriate to describe variants; they are simple enough for video surveillance experts to express their requirements. Yet, they are powerful enough to be liable to static analysis [70]. In particular, the constraints can be analysed as a SAT problem.

An additional challenge is to manage the possible run-time changes of implementation due to context variations (e.g., lighting conditions, changes in the reference scene, etc.). Video surveillance systems have to dynamically adapt to a changing environment. The use of models at run-time is a solution. We are defining adaptation rules corresponding to the dependency constraints between specification elements in one model and software variants in the other [51], [80], [72].

WIMMICS Team

3. Scientific Foundations

3.1. Analyzing and Modeling Users, Communities and their Interactions in a Social Semantic Web Context

We rely on cognitive studies to build models of the system, the user and the interactions between users through the system, in order to support and improve these interactions.

In the short term, following the user modeling technique known as *Personas*, we are interested in these user models that are represented as specific, individual humans. *Personas* are derived from significant behavior patterns (i.e., sets of behavioral variables) elicited from interviews with and observations of users (and sometimes customers) of the future product. Our user models will specialize *Personas* approaches to include aspects appropriate to Web applications. The formalization of these models will rely on ontology-based modeling of users and communities starting with generalist schemas (e.g. FOAF: *Friend of a Friend*). In the longer term we will consider additional extensions of these schemas to capture additional aspects (e.g. emotional states). We will extend current descriptions of relational and emotional aspects in existing variants of the *Personas* technique.

Beyond the individual user models, we propose to rely on social studies to build models of the communities, their vocabularies, activities and protocols in order to identify where and when formal semantics is useful. In the short term we will further develop our method for elaborating collective personas and compare it to the related *collaboration personas* method and to the group modeling methods which are extensions to groups of the classical user modeling techniques dedicated to individuals. We also propose to rely on and adapt participatory sketching and prototyping to support the design of interfaces for visualizing and manipulating representations of collectives. In the longer term we want to focus on studying and modeling mixed representations containing social semantic representations (e.g. folksonomies) and formal semantic representations (e.g. ontologies) and propose operations that allow us to couple them and exchange knowledge between them.

Since we have a background in requirement models, we want to consider in the short term their formalization too in order to support mutual understanding and interoperability between requirements expressed with these heterogeneous models. In a longer term, we believe that argumentation theory can be combined to requirement engineering to improve participant awareness and support decision-making. On the methodological side, we propose to adapt to the design of such systems the incremental formalization approach originally introduced in the context of CSCW (Computer Supported Cooperative Work) and HCI (Human Computer Interaction) communities.

Finally, in the short term, for all the models we identified here we will rely on and evaluate knowledge representation methodologies and theories, in particular ontology-based modeling. In the longer term, additional models of the contexts, devices, processes and mediums will also be formalized and used to support adaptation, proof and explanation and foster acceptation and trust from the users. We specifically target a unified formalization of these contextual aspects to be able to integrate them at any stage of the processing.

3.2. Formalizing and Reasoning on Heterogeneous Semantic Graphs

Our second line of work is to formalize as typed graphs the models identified in the previous section in order for software to exploit them in their processing. The challenge then is two-sided:

- To propose models and formalisms to capture and merge representations of both kinds of semantics (e.g. formal ontologies and social folksonomies). The important point is to allow us to capture those structures precisely and flexibly and yet create as many links as possible between these different objects.

- To propose algorithms (in particular graph-based reasoning) and approaches (e.g. human-computing methods) to process these mixed representations. In particular we are interested in allowing cross-enrichment between them and in exploiting the life cycle and specificities of each one to foster the life-cycles of the others.

While some of these problems are known, for instance in the field of knowledge representation and acquisition (e.g. disambiguation, fuzzy representations, argumentation theory), the Web reopens them with exacerbated difficulties of scale, speed, heterogeneity, and an open-world assumption.

Many approaches emphasize the logical aspect of the problem especially because logics are close to computer languages. We defend that the graph nature of linked data on the Web and the large variety of types of links that compose them call for typed graphs models. We believe the relational dimension is of paramount importance in these representations and we propose to consider all these representations as fragments of a typed graph formalism directly built above the Semantic Web formalisms. Our choice of a graph based programming approach for the semantic and social Web and of a focus on one graph based formalism is also an efficient way to support interoperability, genericity, uniformity and reuse.

ZENITH Project-Team

3. Scientific Foundations

3.1. Data Management

Data management is concerned with the storage, organization, retrieval and manipulation of data of all kinds, from small and simple to very large and complex. It has become a major domain of computer science, with a large international research community and a strong industry. Continuous technology transfer from research to industry has led to the development of powerful DBMSs, now at the heart of any information system, and of advanced data management capabilities in many kinds of software products (application servers, document systems, search engines, directories, etc.).

The fundamental principle behind data management is *data independence*, which enables applications and users to deal with the data at a high conceptual level while ignoring implementation details. The relational model, by resting on a strong theory (set theory and first-order logic) to provide data independence, has revolutionized data management. The major innovation of relational DBMS has been to allow data manipulation through queries expressed in a high-level (declarative) language such as SQL. Queries can then be automatically translated into optimized query plans that take advantage of underlying access methods and indices. Many other advanced capabilities have been made possible by data independence : data and metadata modeling, schema management, consistency through integrity rules and triggers, transaction support, etc.

This data independence principle has also enabled DBMS to continuously integrate new advanced capabilities such as object and XML support and to adapt to all kinds of hardware/software platforms from very small smart devices (smart phone, PDA, smart card, etc.) to very large computers (multiprocessor, cluster, etc.) in distributed environments.

Following the invention of the relational model, research in data management has continued with the elaboration of strong database theory (query languages, schema normalization, complexity of data management algorithms, transaction theory, etc.) and the design and implementation of DBMS. For a long time, the focus was on providing advanced database capabilities with good performance, for both transaction processing and decision support applications. And the main objective was to support all these capabilities within a single DBMS.

The problems of scientific data management (massive scale, complexity and heterogeneity) go well beyond the traditional context of DBMS. To address them, we capitalize on scientific foundations in closely related domains: distributed data management, cloud data management, uncertain data management, metadata integration, data mining and content-based information retrieval.

3.2. Distributed Data Management

To deal with the massive scale of scientific data, we exploit large-scale distributed systems, with the objective of making distribution transparent to the users and applications. Thus, we capitalize on the principles of large-scale distributed systems such as clusters, peer-to-peer (P2P) and cloud, to address issues in data integration, scientific workflows, recommendation, query processing and data analysis.

Data management in distributed systems has been traditionally achieved by distributed database systems which enable users to transparently access and update several databases in a network using a high-level query language (e.g. SQL) [13]. Transparency is achieved through a global schema which hides the local databases' heterogeneity. In its simplest form, a distributed database system is a centralized server that supports a global schema and implements distributed database techniques (query processing, transaction management, consistency management, etc.). This approach has proved effective for applications that can benefit from centralized control and full-fledge database capabilities, e.g. information systems. However, it cannot scale up to more than tens of databases. Data integration systems, e.g. price comparators such as KelKoo, extend the distributed database approach to access data sources on the Internet with a simpler query language in read-only mode.

Parallel database systems extend the distributed database approach to improve performance (transaction throughput or query response time) by exploiting database partitioning using a multiprocessor or cluster system. Although data integration systems and parallel database systems can scale up to hundreds of data sources or database partitions, they still rely on a centralized global schema and strong assumptions about the network.

Scientific workflow management systems (SWfMS) such as Kepler (http://kepler-project.org) and Taverna (http://www.taverna.org.uk) allow scientists to describe and execute complex scientific procedures and activities, by automating data derivation processes, and supporting various functions such as provenance management, queries, reuse, etc. Some workflow activities may access or produce huge amounts of distributed data and demand high performance computing (HPC) environments with highly distributed data sources and computing resources. However, combining SWfMS with HPC to improve throughput and performance remains a difficult challenge. In particular, existing workflow development and computing environments have limited support for data parallelism patterns. Such limitation makes complex the automation and ability to perform efficient parallel execution on large sets of data, which may significantly slow down the execution of a workflow.

In contrast, peer-to-peer (P2P) systems [9] adopt a completely decentralized approach to data sharing. By distributing data storage and processing across autonomous peers in the network, they can scale without the need for powerful servers. Popular examples of P2P systems such as Gnutella and BitTorrent have millions of users sharing petabytes of data over the Internet. Although very useful, these systems are quite simple (e.g. file sharing), support limited functions (e.g. keyword search) and use simple techniques (e.g. resource location by flooding) which have performance problems. To deal with the dynamic behavior of peers that can join and leave the system at any time, they rely on the fact that popular data get massively duplicated.

Initial research on P2P systems has focused on improving the performance of query routing in the unstructured systems which rely on flooding, whereby peers forward messages to their neighbors. This work led to structured solutions based on Distributed Hash Tables (DHT), e.g. CHORD and Pastry, or hybrid solutions with super-peers that index subsets of peers. Another approach is to exploit gossiping protocols, also known as epidemic protocols. Gossiping has been initially proposed to maintain the mutual consistency of replicated data by spreading replica updates to all nodes over the network. It has since been successfully used in P2P networks for data dissemination. Basic gossiping is simple. Each peer has a complete view of the network (i.e. a list of all peers' addresses) and chooses a node at random to spread the request. The main advantage of gossiping is robustness over node failures since, with very high probability, the request is eventually propagated to all nodes in the network. In large P2P networks, however, the basic gossiping model does not scale as maintaining the complete view of the network at each node would generate very heavy communication traffic. A solution to scalable gossiping is by having each peer with only a partial view of the network, e.g. a list of tens of neighbor peers. To gossip a request, a peer chooses at random a peer in its partial view to send it the request. In addition, the peers involved in a gossip exchange their partial views to reflect network changes in their own views. Thus, by continuously refreshing their partial views, nodes can self-organize into randomized overlays which scale up very well.

We claim that a P2P solution is the right solution to support the collaborative nature of scientific applications as it provides scalability, dynamicity, autonomy and decentralized control. Peers can be the participants or organizations involved in collaboration and may share data and applications while keeping full control over their (local) data sources.

But for very-large scale scientific data analysis or to execute very large data-intensive workflow activities (activities that manipulate huge amounts of data), we believe cloud computing (see next section), is the right approach as it can provide virtually infinite computing, storage and networking resources. However, current cloud architectures are proprietary, ad-hoc, and may deprive users of the control of their own data. Thus, we postulate that a hybrid P2P/cloud architecture is more appropriate for scientific data management, by combining the bests of both. In particular, it will enable the clean integration of the users' own computational resources with different clouds.

3.3. Cloud Data Management

Cloud computing encompasses on demand, reliable services provided over the Internet (typically represented as a cloud) with easy access to virtually infinite computing, storage and networking resources. Through very simple Web interfaces and at small incremental cost, users can outsource complex tasks, such as data storage, system administration, or application deployment, to very large data centers operated by cloud providers. Thus, the complexity of managing the software/hardware infrastructure gets shifted from the users' organization to the cloud provider. From a technical point of view, the grand challenge is to support in a cost-effective way the very large scale of the infrastructure which has to manage lots of users and resources with high quality of service.

Cloud customers could move all or part of their information technology (IT) services to the cloud, with the following main benefits:

- **Cost.** The cost for the customer can be greatly reduced since the IT infrastructure does not need to be owned and managed; billing is only based only on resource consumption. For the cloud provider, using a consolidated infrastructure and sharing costs for multiple customers reduces the cost of ownership and operation.

- **Ease of access and use.** The cloud hides the complexity of the IT infrastructure and makes location and distribution transparent. Thus, customers can have access to IT services anytime, and from anywhere with an Internet connection.

- **Quality of Service (QoS).** The operation of the IT infrastructure by a specialized provider that has extensive experience in running very large infrastructures (including its own infrastructure) increases QoS.

- **Elasticity.** The ability to scale resources out, up and down dynamically to accommodate changing conditions is a major advantage. In particular, it makes it easy for customers to deal with sudden increases in loads by simply creating more virtual machines.

However, cloud computing has some drawbacks and not all applications are good candidates for being "cloudified". The major concern is wrt. data security and privacy, and trust in the provider (which may use no so trustful providers to operate). One earlier criticism of cloud computing was that customers get locked in proprietary clouds. It is true that most clouds are proprietary and there are no standards for cloud interoperability. But this is changing with open source cloud software such as Hadoop, an Apache project implementing Google's major cloud services such as Google File System and MapReduce, and Eucalyptus, an open source cloud software infrastructure, which are attracting much interest from research and industry.

There is much more variety in cloud data than in scientific data since there are many different kinds of customers (individuals, SME, large corporations, etc.). However, we can identify common features. Cloud data can be very large, unstructured (e.g. text-based) or semi-structured, and typically append-only (with rare updates). And cloud users and application developers may be in high numbers, but not DBMS experts.

Current cloud data management (NOSQL) solutions typically trade consistency for scalability, simplicity and flexibility. They use a radically different architecture than RDBMS, by exploiting (rather than embedding) a distributed file system such as Google File System (GFS) or Hadoop Distributed File System (HDFS), to store and manage data in a highly fault-tolerant manner. They tend to rely on a more specific data model, e.g. key-value store such Google Bigtable, Hadoop Hbase or Apache CouchDB) with a simple set of operators easy to use from a programming language. For instance, to address the requirements of social network applications, new solutions rely on a graph data model and graph-based operators. User-defined functions also allow for more specific data processing. MapReduce is a good example of generic parallel data processing framework, on top of a distributed file system (GFS or HDFS). It supports a simple data model (sets of (key, value) pairs), which allows user-defined functions (map and reduce). Although quite successful among developers, it is relatively low-level and rigid, leading to custom user code that is hard to maintain and reuse. In Zenith, we exploit or extend these NOSQL technologies to fit our needs for scientific workflow management and scalable data analysis.

3.4. Uncertain Data Management

Data uncertainty is present in many scientific applications. For instance, in the monitoring of plant contamination by INRA teams, sensors generate periodically data which may be uncertain. Instead of ignoring (or correcting) uncertainty, which may generate major errors, we need to manage it rigorously and provide support for querying.

To deal with uncertainty, there are several approaches, e.g. probabilistic, possibilistic, fuzzy logic, etc. The *probabilistic approach* is often used by scientists to model the behavior of their underlying environments. However, in many scientific applications, data management and uncertain query processing are not integrated, i.e. the queries are usually answered using ad-hoc methods after doing manual or semi-automatic statistical treatment on the data which are retrieved from a database. In Zenith, we aim at integrating scientific data management and query processing within one system. This should allow scientists to issue their queries in a query language without thinking about the probabilistic treatment which should be done in background in order to answer the queries. There are two important issues which any PDBMS should address: 1) how to represent a probabilistic database, i.e. data model; 2) how to answer queries using the chosen representation, i.e. query evaluation.

One of the problems on which we focus is *scalable query processing* over uncertain data. A naive solution for evaluating probabilistic queries is to enumerate all possible worlds, i.e. all possible instances of the database, execute the query in each world, and return the possible answers together with their cumulative probabilities. However, this solution can not scale up due to the exponential number of possible worlds which a probabilistic database may have. Thus, the problem is quite challenging, particularly due to exponential number of possibilities that should be considered for evaluating queries. In addition, most of our underlying scientific applications are not centralized; the scientists share part of their data in a *P2P* manner. This distribution of data makes very complicated the processing of probabilistic queries. To develop efficient query processing techniques for distributed scientific applications, we can take advantage of two main distributed technologies: *P2P* and *Cloud*. Our research experience in P2P systems has proved us that we can propose scalable solutions for many data management problems. In addition, we can use the cloud parallel solutions, e.g. MapReduce, to parallelize the task of query processing, when possible, and answer queries of scientists in reasonable execution times. Another challenge for supporting scientific applications is uncertain data integration. In addition to managing the uncertain data for each user, we need to integrate uncertain data from different sources. This requires revisiting traditional data integration in major ways and dealing with the problems of uncertain mediated schema generation and uncertain schema mapping.

3.5. Metadata Integration

Nowdays, scientists can rely on web 2.0 tools to quickly share their data and/or knowledge (e.g. ontologies of the domain knowledge). Therefore, when performing a given study, a scientist would typically need to access and integrate data from many data sources (including public databases). To make high numbers of scientific data sources easily accessible to community members, it is necessary to identifying semantic correspondences between metadata structures or models of the related data sources. The main underlying task is called matching, which is the process of discovering semantic correspondences between metadata structures such as database schema and ontologies. Ontology is a formal and explicit description of a shared conceptualization in term of concepts (i.e., classes, properties and relations). For example, the matching may be used to align gene ontologies or anatomical metadata structures.

To understand a data source content, metadata (data that describe the data) is crucial. Metadata can be initially provided by the data publisher to describe the data structure (e.g. schema), data semantics based on ontologies (that provide a formal representation of the domain knowledge) and other useful information about data provenance (publisher, tools, methods, etc.). Scientific metadata is very heterogeneous, in particular because of the great autonomy of the underlying data sources, which leads to a large variety of models and formats. The high heterogeneity makes the matching problem very challenging. Furthermore, the number of ontologies and their size grow fastly, so does their diversity and heterogeneity. As a result, schema/ontology matching has become a prominent and challenging topic [4].

3.6. Data Mining

Data mining provides methods to discover new and useful patterns from very large sets of data. These patterns may take different forms, depending on the end-user's request, such as:

- **Frequent itemsets and association rules [1]**. In this case, the data is usually a table with a high number of rows and the algorithm extracts correlations between column values. This problem was first motivated by commercial and marketing purposes (*e.g.* discovering frequent correlations between items bought in a shop, which could help selling more). A typical example of frequent itemset from a sensor network in a smart building would say that "in 20% rooms, the door is closed, the room is empty, and lights are on."

- **Frequent sequential pattern extraction.** This problem is very similar to frequent itemset mining, but in this case, the order between events has to be considered. Let us consider the smart-building example again. A frequent sequence, in this case, could say that "in 40% rooms, lights are on at time i, the room is empty at time i+j and the door is closed at time i+j+k". Discovering frequent sequences has become a crucial need in marketing, but also in security (detecting network intrusions for instance) in usage analysis (web usage is one of the main applications) and any domain where data arrive in a specific order (usually given by timestamps).

- **Clustering [12]**. The goal of clustering algorithms is to group together data that have similar characteristics, while ensuring that dissimilar data will not be in the same cluster. In our example of smart buildings, we would find clusters of rooms, where offices will be in one category and copy machine rooms in another one because of their characteristics (hours of people presence, number of times lights are turned on and off, etc.).

One of the main problems for data mining methods recently was to deal with data streams. Actually, data mining methods have first been designed for very large data sets where complex algorithms of artificial intelligence were not able to complete within reasonable time responses because of data size. The problem was thus to find a good trade-off between time response and results relevance. The patterns described above well match this trade-off since they both provide interesting knowledge for data analysts and allow algorithm having good time complexity on the number of records. Itemset mining algorithms, for instance, depend more on the number of columns (for a sensor it would be the number of possible items such as temperature, presence, status of lights, etc.) than the number of lines (number of sensors in the network). However, with the ever growing size of data and their production rate, a new kind of data source has recently emerged as data streams. A data stream is a sequence of events arriving at high rate. By "high rate", we usually admit that traditional data mining methods reach their limits and cannot complete in real-time, given the data size. In order to extract knowledge from such streams, a new trade-off had to be found and the data mining community has investigated approximation methods that could allow maintaining a good quality of results for the above patterns extraction.

For scientific data, data mining now has to deal with new and challenging characteristics. First, scientific data is often associated to a level of uncertainty (typically, sensed values have to be associated to the probability that this value is correct or not). Second, scientific data might be extremely large and need cloud computing solutions for their storage and analysis. Eventually, we will have to deal with high dimension and heterogeneous data.

3.7. Content-based Information Retrieval

Today's technologies for searching information in scientific data mainly rely on relational DBMS or text based indexing methods. However, content-based information retrieval has progressed much in the last decade and is now considered as one of the most promising for future search engines. Rather than restricting search to the use of metadata, content-based methods attempt to index, search and browse digital objects by means of signatures describing their actual content. Such methods have been intensively studied in the multimedia community to allow searching the massive amount or raw multimedia documents created every day (e.g. 99% of web data are audio-visual content with very sparse metadata). Successful and scalable content-based

methods have been proposed for searching objects in large image collections or detecting copies in huge video archives. Besides multimedia contents, content-based information retrieval methods recently started to be studied on more diverse data such as medical images, 3D models or even molecular data. Potential applications in scientific data management are numerous. First of all, to allow searching the huge collections of scientific images (earth observation, medical images, botanical images, biology images, etc.) but also to browse large datasets of experimental data (e.g. multisensor data, molecular data or instrumental data). Despite recent progress, scalability remains a major issue, involving complex algorithms (such as similarity search, clustering or supervised retrieval), in high dimensional spaces (up to millions of dimensions) with complex metrics (Lp, Kernels, sets intersections, edit distances, etc.). Most of these algorithms have linear, quadratic or even cubic complexities so that their use at large scale is not affordable without consistent breakthrough. In Zenith, we plan to investigate the following challenges:

- **High-dimensional similarity search**. Whereas many indexing methods were designed in the last 20 years to retrieve efficiently multidimensional data with relatively small dimensions, high-dimensional data have been more challenging due to the well-known dimensionality curse. Only recently have some methods appeared that allow approximate Nearest Neighbors queries in sub-linear time, in particular, Locality Sensitive Hashing methods which offer new theoretical insights in high-dimensional Euclidean spaces and proved the interest of random projections. But there are still some challenging issues that need to be solved including efficient similarity search in any kernel or metric spaces, efficient construction of knn-graphs or relational similarity queries.

- **Large-scale supervised retrieval**. Supervised retrieval aims at retrieving relevant objects in a dataset by providing some positive and/or negative training samples. To solve such task, there has been a focused interest on using Support Vector Machines (SVM) that offer the possibility to construct generalized, non-linear predictors in high-dimensional spaces using small training sets. The prediction time complexity of these methods is usually linear in dataset size. Allowing hyperplane similarity queries in sub-linear time is for example a challenging research issue. A symmetric problem in supervised retrieval consists in retrieving the most relevant object categories that might contain a given query object, providing huge labeled datasets (up to millions of classes and billions of objects) and very few objects per category (from 1 to 100 objects). SVM methods that are formulated as quadratic programming with cubic training time complexity and quadratic space complexity are clearly not usable. Promising solutions to such problems include hybrid supervised-unsupervised methods and supervised hashing methods.

- **P2P content-based retrieval**. Content-based P2P retrieval methods appeared recently as a promising solution to manage masses of data distributed over large social networks, particularly when the data cannot be centralized for privacy or cost reasons (which is often the case in scientific social networks, e.g. botanist social networks). However, current methods are limited to very simple similarity search paradigms. In Zenith, we will consider more advanced P2P content-based retrieval and mining methods such as k-nn graphs construction, large-scale supervised retrieval or multi-source clustering.