RESEARCH CENTER
Bordeaux - Sud-Ouest

FIELD

Activity Report 2013

Section Scientific Foundations

Edition: 2014-03-19

LFANT Project-Team

3. Research Program

3.1. Number fields, class groups and other invariants

Participants: Bill Allombert, Athanasios Angelakis, Karim Belabas, Julio Brau, Jean-Paul Cerri, Henri Cohen, Jean-Marc Couveignes, Andreas Enge, Pierre Lezowski, Nicolas Mascot, Aurel Page.

Modern number theory has been introduced in the second half of the 19th century by Dedekind, Kummer, Kronecker, Weber and others, motivated by Fermat's conjecture: There is no non-trivial solution in integers to the equation $x^n + y^n = z^n$ for $n \geqslant 3$. For recent textbooks, see [5]. Kummer's idea for solving Fermat's problem was to rewrite the equation as $(x+y)(x+\zeta y)(x+\zeta^2 y)\cdots(x+\zeta^{n-1}y) = z^n$ for a primitive n-th root of unity ζ, which seems to imply that each factor on the left hand side is an n-th power, from which a contradiction can be derived.

The solution requires to augment the integers by *algebraic numbers*, that are roots of polynomials in $\mathbb{Z}[X]$. For instance, ζ is a root of $X^n - 1$, $\sqrt[3]{2}$ is a root of $X^3 - 2$ and $\frac{\sqrt{3}}{5}$ is a root of $25X^2 - 3$. A *number field* consists of the rationals to which have been added finitely many algebraic numbers together with their sums, differences, products and quotients. It turns out that actually one generator suffices, and any number field K is isomorphic to $\mathbb{Q}[X]/(f(X))$, where $f(X)$ is the minimal polynomial of the generator. Of special interest are *algebraic integers*, "numbers without denominators", that are roots of a monic polynomial. For instance, ζ and $\sqrt[3]{2}$ are integers, while $\frac{\sqrt{3}}{5}$ is not. The *ring of integers* of K is denoted by $\mathcal{O}_K$; it plays the same role in K as $\mathbb{Z}$ in $\mathbb{Q}$.

Unfortunately, elements in $\mathcal{O}_K$ may factor in different ways, which invalidates Kummer's argumentation. Unique factorisation may be recovered by switching to *ideals*, subsets of $\mathcal{O}_K$ that are closed under addition and under multiplication by elements of $\mathcal{O}_K$. In $\mathbb{Z}$, for instance, any ideal is *principal*, that is, generated by one element, so that ideals and numbers are essentially the same. In particular, the unique factorisation of ideals then implies the unique factorisation of numbers. In general, this is not the case, and the *class group* Cl_K of ideals of $\mathcal{O}_K$ modulo principal ideals and its *class number* $h_K = |\,\mathrm{Cl}_K\,|$ measure how far $\mathcal{O}_K$ is from behaving like $\mathbb{Z}$.

Using ideals introduces the additional difficulty of having to deal with *units*, the invertible elements of $\mathcal{O}_K$: Even when $h_K = 1$, a factorisation of ideals does not immediately yield a factorisation of numbers, since ideal generators are only defined up to units. For instance, the ideal factorisation $(6) = (2)\cdot(3)$ corresponds to the two factorisations $6 = 2\cdot 3$ and $6 = (-2)\cdot(-3)$. While in $\mathbb{Z}$, the only units are 1 and -1, the unit structure in general is that of a finitely generated $\mathbb{Z}$-module, whose generators are the *fundamental units*. The *regulator* R_K measures the "size" of the fundamental units as the volume of an associated lattice.

One of the main concerns of algorithmic algebraic number theory is to explicitly compute these invariants (Cl_K and h_K, fundamental units and R_K), as well as to provide the data allowing to efficiently compute with numbers and ideals of $\mathcal{O}_K$; see [32] for a recent account.

The *analytic class number formula* links the invariants h_K and R_K (unfortunately, only their product) to the ζ-function of K, $\zeta_K(s) := \prod_{\mathfrak{p} \text{ prime ideal of } \mathcal{O}_K} (1 - \mathrm{N}\,\mathfrak{p}^{-s})^{-1}$, which is meaningful when $\Re(s) > 1$, but which may be extended to arbitrary complex $s \neq 1$. Introducing characters on the class group yields a generalisation of ζ- to L-functions. The *generalised Riemann hypothesis (GRH)*, which remains unproved even over the rationals, states that any such L-function does not vanish in the right half-plane $\Re(s) > 1/2$. The validity of the GRH has a dramatic impact on the performance of number theoretic algorithms. For instance, under GRH, the class group admits a system of generators of polynomial size; without GRH, only exponential bounds are known. Consequently, an algorithm to compute Cl_K via generators and relations (currently the only viable practical approach) either has to assume that GRH is true or immediately becomes exponential.

When $h_K = 1$ the number field K may be norm-Euclidean, endowing $\mathcal{O}_K$ with a Euclidean division algorithm. This question leads to the notions of the Euclidean minimum and spectrum of K, and another task in algorithmic number theory is to compute explicitly this minimum and the upper part of this spectrum, yielding for instance generalised Euclidean gcd algorithms.

3.2. Function fields, algebraic curves and cryptology

Participants: Karim Belabas, Julio Brau, Jean-Marc Couveignes, Andreas Enge, Nicolas Mascot, Jérôme Milan, Damien Robert, Vincent Verneuil.

Algebraic curves over finite fields are used to build the currently most competitive public key cryptosystems. Such a curve is given by a bivariate equation $\mathcal{C}(X, Y) = 0$ with coefficients in a finite field $\mathbb{F}_q$. The main classes of curves that are interesting from a cryptographic perspective are *elliptic curves* of equation $\mathcal{C} = Y^2 - (X^3 + aX + b)$ and *hyperelliptic curves* of equation $\mathcal{C} = Y^2 - (X^{2g+1} + \cdots)$ with $g \geqslant 2$.

The cryptosystem is implemented in an associated finite abelian group, the *Jacobian* $\mathrm{Jac}_\mathcal{C}$. Using the language of function fields exhibits a close analogy to the number fields discussed in the previous section. Let $\mathbb{F}_q(X)$ (the analogue of $\mathbb{Q}$) be the *rational function field* with subring $\mathbb{F}_q[X]$ (which is principal just as $\mathbb{Z}$). The *function field* of $\mathcal{C}$ is $K_\mathcal{C} = \mathbb{F}_q(X)[Y]/(\mathcal{C})$; it contains the *coordinate ring* $\mathcal{O}_\mathcal{C} = \mathbb{F}_q[X, Y]/(\mathcal{C})$. Definitions and properties carry over from the number field case $K/\mathbb{Q}$ to the function field extension $K_\mathcal{C}/\mathbb{F}_q(X)$. The Jacobian $\mathrm{Jac}_\mathcal{C}$ is the divisor class group of $K_\mathcal{C}$, which is an extension of (and for the curves used in cryptography usually equals) the ideal class group of $\mathcal{O}_\mathcal{C}$.

The size of the Jacobian group, the main security parameter of the cryptosystem, is given by an L-function. The GRH for function fields, which has been proved by Weil, yields the Hasse–Weil bound $(\sqrt{q} - 1)^{2g} \leqslant |\mathrm{Jac}_\mathcal{C}| \leqslant (\sqrt{q} + 1)^{2g}$, or $|\mathrm{Jac}_\mathcal{C}| \approx q^g$, where the *genus* g is an invariant of the curve that correlates with the degree of its equation. For instance, the genus of an elliptic curve is 1, that of a hyperelliptic one is $\frac{\deg_X \mathcal{C} - 1}{2}$. An important algorithmic question is to compute the exact cardinality of the Jacobian.

The security of the cryptosystem requires more precisely that the *discrete logarithm problem* (DLP) be difficult in the underlying group; that is, given elements D_1 and $D_2 = xD_1$ of $\mathrm{Jac}_\mathcal{C}$, it must be difficult to determine x. Computing x corresponds in fact to computing $\mathrm{Jac}_\mathcal{C}$ explicitly with an isomorphism to an abstract product of finite cyclic groups; in this sense, the DLP amounts to computing the class group in the function field setting.

For any integer n, the *Weil pairing* e_n on $\mathcal{C}$ is a function that takes as input two elements of order n of $\mathrm{Jac}_\mathcal{C}$ and maps them into the multiplicative group of a finite field extension $\mathbb{F}_{q^k}$ with $k = k(n)$ depending on n. It is bilinear in both its arguments, which allows to transport the DLP from a curve into a finite field, where it is potentially easier to solve. The *Tate-Lichtenbaum pairing*, that is more difficult to define, but more efficient to implement, has similar properties. From a constructive point of view, the last few years have seen a wealth of cryptosystems with attractive novel properties relying on pairings.

For a random curve, the parameter k usually becomes so big that the result of a pairing cannot even be output any more. One of the major algorithmic problems related to pairings is thus the construction of curves with a given, smallish k.

3.3. Complex multiplication

Participants: Karim Belabas, Henri Cohen, Jean-Marc Couveignes, Andreas Enge, Nicolas Mascot, Enea Milio, Aurel Page, Damien Robert.

Complex multiplication provides a link between number fields and algebraic curves; for a concise introduction in the elliptic curve case, see [39], for more background material, [37]. In fact, for most curves $\mathcal{C}$ over a finite field, the endomorphism ring of $\mathrm{Jac}_\mathcal{C}$, which determines its L-function and thus its cardinality, is an order in a special kind of number field K, called *CM field*. The CM field of an elliptic curve is an imaginary-quadratic field $\mathbb{Q}(\sqrt{D})$ with $D < 0$, that of a hyperelliptic curve of genus g is an imaginary-quadratic extension of a totally real number field of degree g. Deuring's lifting theorem ensures that $\mathcal{C}$ is the reduction modulo some prime of a curve with the same endomorphism ring, but defined over the *Hilbert class field* H_K of K.

Algebraically, H_K is defined as the maximal unramified abelian extension of K; the Galois group of H_K/K is then precisely the class group Cl_K. A number field extension H/K is called *Galois* if $H \simeq K[X]/(f)$ and H contains all complex roots of f. For instance, $\mathbb{Q}(\sqrt{2})$ is Galois since it contains not only $\sqrt{2}$, but also the second root $-\sqrt{2}$ of X^2-2, whereas $\mathbb{Q}(\sqrt[3]{2})$ is not Galois, since it does not contain the root $e^{2\pi i/3}\sqrt[3]{2}$ of X^3-2. The *Galois group* $\mathrm{Gal}_{H/K}$ is the group of automorphisms of H that fix K; it permutes the roots of f. Finally, an *abelian* extension is a Galois extension with abelian Galois group.

Analytically, in the elliptic case H_K may be obtained by adjoining to K the *singular value* $j(\tau)$ for a complex valued, so-called *modular* function j in some $\tau \in \mathcal{O}_K$; the correspondence between $\mathrm{Gal}_{H/K}$ and Cl_K allows to obtain the different roots of the minimal polynomial f of $j(\tau)$ and finally f itself. A similar, more involved construction can be used for hyperelliptic curves. This direct application of complex multiplication yields algebraic curves whose L-functions are known beforehand; in particular, it is the only possible way of obtaining ordinary curves for pairing-based cryptosystems.

The same theory can be used to develop algorithms that, given an arbitrary curve over a finite field, compute its L-function.

A generalisation is provided by *ray class fields*; these are still abelian, but allow for some well-controlled ramification. The tools for explicitly constructing such class fields are similar to those used for Hilbert class fields.

ALEA Project-Team

3. Research Program

3.1. Research Program

This idea of analyzing nature systems and transferring the underlying principles into stochastic algorithms and technical implementations is one of the central component of the ALEA team project. Adapting nature mechanisms and biological capabilities clearly provides a better understanding of the real processes, and it also improves the performance and the power of engineers devices. Our project is centered on both the understanding of biological processes in terms of mathematical, physical and chemical models, and on the other hand, on the use of these biology inspired stochastic algorithms to solve complex engineering problems.

There is a huge series of virtual interfaces, robotic devices, numerical schemes and stochastic algorithms which were invented mimicking biological processes or simulating natural mechanisms. The terminology *"mimicking or simulating"* doesn't really mean to find an exact copy of natural processes, but *to elaborate the mathematical principles so that they can be abstracted from the original biological or physical model.* In our context, the whole series of evolutionary type principles discussed in previous sections can be abstracted into only three different and natural classes of stochastic algorithms, depending on the nature of the biology-inspired interaction mechanism used in the stochastic evolution model. These three stochastic search models are listed below :

1) *Branching and interacting particle systems (birth and death chains, spatial branching processes, mean-field interaction between generations):*

The first generation of adaptive branching-selection algorithms is very often built on the same genetic type paradigm: When exploring a state space with many particles, we duplicate better fitted individuals at the expense of light particles with poor fitness die. From a computational point of view, we generate a large number of random problem solvers. Each one is then rated according to a fitness or performance function defined by the developer. Mimicking natural selection, an evolutionary algorithm selects the best solvers in each generation and breeds them.

2) *Reinforced random walks and self-interacting chains (reinforced learning strategies, interaction processes with respect to the occupation measure of the past visited sites):*

This type of reinforcement is observed frequently in nature and society, where "beneficial" interactions with the past history tend to be repeated. A new class of historical mean field type interpretation models of reinforced processes were developed by the team project leader in a pair of articles [32], [31]. Self interaction gives the opportunity to build new stochastic search algorithms with the ability to, in a sense, re-initialized their exploration from the past, re-starting from some better fitted initial value already met in the past [33], [34].

3) *Random tree based stochastic exploration models (coalescent and genealogical tree search explorations techniques on path space):*

The last generation of stochastic random tree models is concerned with biology-inspired algorithms on paths and excursions spaces. These genealogical adaptive search algorithms coincide with genetic type particle models in excursion spaces. They have been applied with success in generating the excursion distributions of Markov processes evolving in critical and rare event regimes, as well as in path estimation and related smoothing problems arising in advanced signal processing (cf. [29] and references therein). We underline the fact that the complete mathematical analysis of these random tree models, including their long time behavior, their propagations of chaos properties, as well as their combinatorial structures are far from being completed. This class of genealogical tree based models has been introduced in [30] for solving smoothing problems and more generally Feynman-Kac semigroups on path spaces, see also [28], [29], and references therein.

BACCHUS Team

3. Research Program

3.1. Numerical schemes for fluid mechanics

Participants: Rémi Abgrall, Mario Ricchiuto, Dante de Santis, Pietro Marco Congedo, Cécile Dobrzynzki, Héloïse Beaugendre, Pierre-Henri Maire, Luc Mieussens, Philippe Bonneton, Gérard Vignoles.

A large number of engineering problems involve fluid mechanics. They may involve the coupling of one or more physical models. An example is provided by aeroelastic problems, which have been studied in details by other Inria teams. Another example is given by flows in pipelines where the fluid (a mixture of air–water–gas) does not have well-known physical properties, and there are even more exotic situations. In some occasions, one needs specific numerical tools to take into account *e.g.* a fluids' exotic equation of state, or a the influence of small flow scales in a macro-/meso-scopic flow model, etc. Efficient schemes are needed in unsteady flows where the amount of required computational resources becomes huge. Another situation where specific tools are needed is when one is interested in very specific physical quantities, such as *e.g.* the lift and drag of an airfoil, or the boundary of the area flooded by a Tsunami.

In these situations, commercial tools can only provide a crude answer. These codes, while allowing users to simulate a lot of different flow types, and "always" providing an answer, give results often of poor quality. This is mainly due to their general purpose character, and on the fact that the numerical technology implemented in these codes is not the most recent. To give a few examples, consider the noise generated by wake vortices in supersonic flows (external aerodynamics/aeroacoustics), or the direct simulation of a 3D compressible mixing layer in a complex geometry (as in combustion chambers). Up to our knowledge, due to the very different temporal and physical scales that need to be captured, a direct simulation of these phenomena is not in the reach of the most recent technologies because the numerical resources required are currently unavailable. *We need to invent specific algorithms for this purpose.*

Our goal is to develop more accurate and more efficient schemes that can adapt to modern computer architectures, and allow the efficient simulation of complex real life flows.

We develop a class of numerical schemes, known in literature as Residual Distribution schemes, *specifically tailored to unstructured and hybrid meshes*. They have the most possible compact stencil that is compatible with the expected order of accuracy. This *accuracy is at least of second order, and it can go up to any order of accuracy, even though fourth order is considered for practical applications.* Since the stencil is compact, the implementation on parallel machines becomes simple. These schemes are very flexible in nature, which is so far one of the most important advantage over other techniques. This feature has allowed us to adapt the schemes to the requirements of different physical situations (*e.g.* different formulations allow either en efficient explicit time advancement for problems involving small time-scales, or a fully implicit space-time variant which is unconditionally stable and allows to handle stiff problems where only the large time scales are relevant). This flexibility has also enabled to devise a variant using the same data structure of the popular Discontinuous Galerkin schemes, which are also part of our scientific focus.

The compactness of the second order version of the schemes enables us to use efficiently the high performance parallel linear algebra tools developed by the team. However, the high order versions of these schemes, which are under development, require modifications to these tools taking into account the nature of the data structure used to reach higher orders of accuracy. This leads to new scientific problems at the border between numerical analysis and computer science. In parallel to these fundamental aspects, we also work on adapting more classical numerical tools to complex physical problems such as those encountered in interface flows, turbulent or multiphase flows, geophysical flows, and material science. A particular attention has been devoted to the implementation of complex thermodynamic models permitting to simulate several classes of fluids and to take into account real-gas effects and some exotic phenomenon, such as rarefaction shock waves.

Within these applications, a strong effort has been made in developing more predictive tools for both multiphase compressible flows and non-hydrostatic free surface flows.

Concerning multiphase flows, several advancements have been performed, i.e. considering a more complete systems of equations including viscosity, working on the thermodynamic modeling of complex fluids, and developing stochastic methods for uncertainty quantification in compressible flows. Concerning depth averaged free surface flow modeling, on one hand we have shown the advantages of the use of the compact schemes we develop for hydrostatic shallow water models. On the other, we have shown ho to extend our approach to non-hydrostatic Boussinesq modeling, including wave dispersion, and wave breaking effects.
We expect to be able to demonstrate the potential of our developments on applications ranging from the the reproduction of the complex multidimensional interactions between tidal waves and estuaries, to the unsteady aerodynamics and aeroacoustics associated to both external and internal compressible flows, and the behavior of complex materials. This will be achieved by means of a multi-disciplinary effort involving our research on residual discretizations schemes, the parallel advances in algebraic solvers and partitioners, and the strong interactions with specialists in computer science, scientific computing, physics, mechanics, and mathematical modeling.
Concerning the software platforms, our research in numerical algorithms has led to the development of the `RealfluiDS` platform which is described in section 5.3 , and to the `SLOWS` (Shallow-water fLOWS) code for free surface flows, described in sections 5.9 . Simultaneously, we have contributed to the advancement of the new, object oriented, parallel finite elements library `AeroSol`, described in section 5.1 , which is destined to replace the existing codes and become the team's CFD kernel.

New software developments are under way in the field of complex materials modeling. These developments are performed in the code in the solver `COCA` (CodeOxydationCompositesAutocicatrisants) for the simulation of the self-healing process in composite materials. These developments will be described in section 5.2 .

This work is supported by the EU-Strep IDIHOM, various research contracts and in part by the `ANEMOS` project and the ANR-Emergence `RealFluids` grant. A large part of the team also profited of the `ADDECCO` ERC grant.

3.2. Numerical schemes for Uncertainty quantification and robust optimization

Participants: Rémi Abgrall, Pietro Marco Congedo, Gianluca Geraci, Mario Ricchiuto, Maria Giovanna Rodio, Francesca Fusi, Julie Tryoen, Kunkun Tang.

Another topic of interest is the quantification of uncertainties in non linear problems. In many applications, the physical model is not known accurately. The typical example is that of turbulence models in aeronautics. These models all depend on a number of parameters which can radically change the output of the simulation. Being impossible to lump the large number of temporal and spatial scales of a turbulent flow in a few model parameters, these values are often calibrated to quantitatively reproduce a certain range of effects observed experimentally. A similar situation is encountered in many applications such as real gas or multiphase flows, where the equation of state form suffer from uncertainties, and free surface flows with sediment transport, where often both the hydrodynamic model and the sediment transport model depend on several parameters, and my have more than one formal expression.

This type of uncertainty, called *epistemic*, is associated with a lack of knowledge and could be reduced by further experiments and investigation. Instead, another type of uncertainty, called *aleatory*, is related to the intrinsec aleatory quality of a physical measure and can not be reduced. The dependency of the numerical simulation from these uncertainties can be studied by propagation of chaos techniques such as those developed during the recent years via polynomial chaos techniques. Different implementations exists, depending whether the method is intrusive or not. The accuracy of these methods is still a matter of research, as well how they can handle an as large as possible number of uncertainties or their versatility with respect to the structure of the random variable pdfs.

Our objective is to develop some non-intrusive and semi-intrusive methods, trying to define an unified framework for obtained a reliable and accurate numerical solution at a moderate computational cost. This work have produced a large number of publications on peer-reviewed journal. Concerning the class of intrusive methods, we are developing an unified scheme in the coupled physical/stochastic space based on a multi-resolution framework. Here, the idea is to build a framework for being capable to refine a discontinuity in both stochastic and deterministic mesh. We are extending this class of methods to complex models in CFD, such as in multiphase flows. Concerning the non-intrusive methods, we are working on several methods for treating the following problems : handling a large number of uncertainties, treating high-order statistical decomposition (variance, skewness and kurtosis), and solving efficiently inverse problems.

We have used these methods to several ends : either to have highly accurate quantitative reconstruction of a simulation output's variation over a complex space of parameter variations to study a given model (uncertainty propagation), or as a means of comparing different model's variability to certain parameters thus assessing their robustness (model robustness), or as a tool to compare different numerical implementation (schemes and codes) of a similar model to assess simultaneously the robustness of the numerics and the universality of the trends of the statistics and of the sensitivity measures (robust cross-validation). Moreover, we rebuild statistically some input parameters relying on some experimental measures of the output, thus solving an inverse problem.

The developed methods and tools have been applied to several applications of interest : real-gas effects, multi-phase flows, cavitation, aerospace applications and geophysical flows.

Concerning robust optimization, we focus on problems with high dimensional representation of stochastic inputs, that can be computationally prohibitive. In fact, for a robust design, statistics of the fitness functions are also important, then uncertainty quantification (UQ) becomes the predom- inant issue to handle if a large number of uncertainties is taken into account. Several methods are proposed in literature to consider high dimension stochastic problem but their accuracy on realistic problems where highly non-linear effects could exist is not proven at all. We developed several efficient global strategies for robust optimization: the first class of method is based on the extension of simplex stochastic collocation to the optimization space, the second one consists in hybrid strategies using ANOVA decomposition.

These developments and computations are performed in the platform `RobUQ`, which includes the most part of methods developed in the Team.

This part of our activities is supported by the ERC grant `ADDECCO`, the ANR-MN project `UFO` and the associated team `AQUARIUS`.

3.3. Meshes and scalable discrete data structures

Participants: Cécile Dobrzynski, Sébastien Fourestier, Algiane Froehly, Cédric Lachat, François Pellegrini.

3.3.1. Adaptive dynamic mesh partitioning

Many simulations which model the evolution of a given phenomenon along with time (turbulence and unsteady flows, for instance) need to re-mesh some portions of the problem graph in order to capture more accurately the properties of the phenomenon in areas of interest. This re-meshing is performed according to criteria which are closely linked to the undergoing computation and can involve large mesh modifications: while elements are created in critical areas, some may be merged in areas where the phenomenon is no longer critical.

Performing such re-meshing in parallel creates additional problems. In particular, splitting an element which is located on the frontier between several processors is not an easy task, because deciding when splitting some element, and defining the direction along which to split it so as to preserve numerical stability most, require shared knowledge which is not available in distributed memory architectures. Ad-hoc data structures and algorithms have to be devised so as to achieve these goals without resorting to extra communication and synchronization which would impact the running speed of the simulation.

Most of the works on parallel mesh adaptation attempt to parallelize in some way all the mesh operations: edge swap, edge split, point insertion, etc. It implies deep modifications in the (re)mesher and often leads to bad performance in term of CPU time. An other work [74] proposes to base the parallel re-meshing on existing mesher and load balancing to be able to modify the elements located on the frontier between several processors.

In addition, the preservation of load balance in the re-meshed simulation requires dynamic redistribution of mesh data across processing elements. Several dynamic repartitioning methods have been proposed in the literature [75], [73], which rely on diffusion-like algorithms and the solving of flow problems to minimize the amount of data to be exchanged between processors. However, integrating such algorithms into a global framework for handling adaptive meshes in parallel has yet to be done.

The path that we are following bases on the decomposition of the areas to remesh into boules that can be processed concurrently, each by a sequential remesher. It requires to devise scalable algorithms for building such boules, scheduling them on as many processors as possible, reconstructing the remeshed mesh and redistributing its data. This research started within the context of the PhD of Cédric Lachat, funded by a CORDI grant of EPI PUMAS and is continued thanks to a funding by ADT grant El Gaucho.

3.3.2. Graph partitioning and static mapping

Unlike their predecessors of two decades ago, today's very large parallel architectures can no longer implement a uniform memory model. They are based on a hierarchical structure, in which cores are assembled into chips, chips are assembled into boards, boards are assembled into cabinets and cabinets are interconnected through high speed, low latency communication networks. On these systems, communication is non uniform: communicating with cores located on the same chip is cheaper than with cores on other boards, and much cheaper than with cores located in other cabinets. The advent of these massively parallel, non uniform machines impacts the design of the software to be executed on them, both for applications and for service tools. It is in particular the case for the software whose task is to balance workload across the cores of these architectures.

A common method for task allocation is to use graph partitioning tools. The elementary computations to perform are represented by vertices and their dependencies by edges linking two vertices that need to share some piece of data. Finding good solutions to the workload distribution problem amounts to computing partitions with small vertex or edge cuts and that balance evenly the weights of the graph parts. Yet, computing efficient partitions for non uniform architectures requires to take into account the topology of the target architecture. When processes are assumed to coexist simultaneously for all the duration of the program, this generalized optimization problem is called mapping. In this problem, the communication cost function to minimize incorporates architecture-dependent, locality improving terms, such as the dilation of each edge (that is, by how much it is "stretched" across the graph representing the target architecture), which is sometimes also expressed as some "hop metric". A mapping is called static if it is computed prior to the execution of the program and is never modified at run-time.

The sequential Scotch tool being developed within the BACCHUS team (see Section 5.8) was able to perform static mapping since its first version, in 1994, but this feature was not widely known nor used by the community. With the increasing need to map very large problem graphs onto very large and strongly non uniform parallel machines, there is an increasing demand for parallel static mapping tools. Since, in the context of dynamic repartitioning, parallel mapping software will have to run on their target architectures, parallel mapping and remapping algorithms suitable for efficient execution on such heterogeneous architectures have to be investigated. This leads to solve three interwoven challenges:

- scalability: such algorithms must be able to map graphs of more than a billion vertices onto target architectures comprising millions of cores;
- heterogeneity: not only do these algorithms must take into account the topology of the target architecture they map graphs onto, but they also have themselves to run efficiently on these very architectures;

- asynchronicity: most parallel partitioning algorithms use collective communication primitives, that is, some form of heavy synchronization. With the advent of machines having several millions of cores, and in spite of the continuous improvement of communication subsystems, the demand for more asynchronicity in parallel algorithms is likely to increase.

This research was mainly carried out within the context of the PhD of Sébastien Fourestier, who defended on June.

CAGIRE Team

3. Research Program

3.1. Computational fluid mechanics: resolving versus modelling small scales of turbulence

A typical continuous solution of the Navier Stokes equations is governed by a spectrum of time and space scales. The broadness of that spectrum is directly controlled by the Reynolds number defined as the ratio between the inertial forces and the viscous forces. This number is quite helpful to determine if the flow is turbulent or not. In the former case, it indicates the range of scales of fluctuations that are present in the flow under study. Typically, for instance for the velocity field, the ratio between the largest scale (the integral length scale) to the smallest one (Kolmogorov scale) scales as $Re^{3/4}$ per dimension. In addition, for internal flows, the viscous effects near the solid walls yield a scaling proportional to Re per dimension. The smallest scales may have a certain effect on the largest ones which implies that an accurate framework for the computation of flows must take into account all these scales. This can be achieved either by solving directly the Navier-Stokes equations (Direct numerical simulations or DNS) or by first applying a time filtering (Reynolds Average Navier-Stokes or RANS) or a spatial filtering operator to the Navier-Stokes equations (large-eddy simulations or LES). The new terms brought about by the filtering operator have to be modelled. From a computational point of view, the RANS approach is the less demanding, which explains why historically it has been the workhorse in both the academic and the industrial sectors. Although it has permitted quite a substantive progress in the understanding of various phenomena such as turbulent combustion or heat transfer, its inability to provide a time-dependent information has led to promote in the last decade the recourse to either LES or DNS. By simulating the large scale structures while modelling the smallest ones supposed to be more isotropic, LES proved to be quite a step through that permits to fully take advantage of the increasing power of computers to study complex flow configurations. In the same time, DNS was progressively applied to geometries of increasing complexity (channel flows, jets, turbulent premixed flames), and proved to be a formidable tool that permits **(i)** to improve our knowledge of turbulent flows and **(ii)** to test (i.e. validate or invalidate) and improve the numerous modelling hypotheses inherently associated to the RANS and LES approaches. From a numerical point of view, if the steady nature of the RANS equations allows to perform iterative convergence on finer and finer meshes, this is no longer possible for LES or DNS which are time-dependent. It is therefore necessary to develop high accuracy schemes in such frameworks. Considering that the Reynolds number in an engine combustion chamber is significantly larger than 10000, a direct numerical simulation of the whole flow domain is not conceivable on a routine basis but the simulation of generic flows which feature some of the phenomena present in a combustion chamber is accessible considering the recent progresses in High Performance Computing (HPC). Along these lines, our objective is to develop a DNS tool to simulate a jet in crossflow configuration which is the generic flow of an aeronautical combustion chamber as far as its effusion cooling is concerned.

3.2. Computational fluid mechanics: numerical methods

All the methods we describe are mesh-based methods: the computational domain is divided into *cells*, that have an elementary shape: triangle and quadrangle in two dimensions, and tetrahedra, hexahedra, pyramids, and prism in three dimensions. If the cells are only regular hexahedra, the mesh is said to be *structured*. Otherwise, it is said to be unstructured. If the mesh is composed of more than one sort of elementary shape, the mesh is said to be *hybrid*.

The basic numerical model for the computation of internal flows is based on the Navier-Stokes equations. For fifty years, many sorts of numerical approximation have been tried for this sort of system: finite differences, finite volumes, and finite elements.

The finite differences have met a great success for some equations, but for the approximation of fluid mechanics, they suffer from two drawbacks. First, structured meshes must be used. This drawback can be very limiting in the context of internal aerodynamics, in which the geometries can be very complex. The other problem is that finite difference schemes do not include any upwinding process, which is essential for convection dominated flows.

The finite volumes methods have imposed themselves in the last thirty years in the context of aerodynamic. They intrinsically contain an upwinding mechanism, so that they are naturally stable for linear as much as for nonlinear convective flows. The extension to diffusive flows has been done in [10]. Whereas the extension to second order with the MUSCL method is widely spread, the extension to higher order has always been a strong drawback of finite volumes methods. For such an extension, reconstruction methods have been developed (ENO, WENO). Nevertheless, these methods need to use a stencil that increases quickly with the order, which induces problems for the parallelisation and the efficiency of the implementation. Another natural extension of finite volume methods are the so-called discontinuous Galerkin methods. These methods are based on the Galerkin' idea of projecting the weak formulation of the equations on a finite dimensional space. But on the contrary to the conforming finite elements method, the approximation space is composed of functions that are continuous (typically: polynomials) inside each cell, but that are discontinuous on the sides. The discontinuous Galerkin methods are currently very popular, because they can be used with many sort of partial differential equations. Moreover, the fact that the approximation is discontinuous allows to use modern mesh adaptation (hanging nodes, which appear in non conforming mesh adaptation), and adaptive order, in which the high order is used only where the solution is smooth.

Discontinuous Galerkin methods where introduced by Reed and Hill [32] and first studied by Lesaint and Raviart [25]. The extension to the Euler system with explicit time integration was mainly led by Shu, Cockburn and their collaborators. The steps of time integration and slope limiting were similar to high order ENO schemes, whereas specific constraints given by the finite elements nature of the scheme were progressively solved, for scalar conservation laws [14], [13], one dimensional systems [12], multidimensional scalar conservation laws [11], and multidimensional systems [15]. For the same system, we can also cite the work of [17], [23], which is slightly different: the stabilisation is made by adding a nonlinear stabilisation term, and the time integration is implicit. Then, the extension to the compressible Navier-Stokes system was made by Bassi and Rebay [9], first by a mixed type finite element method, and then simplified by means of lifting operators. The extension to the $k-\omega$ RANS system was made in [8]. Another type of discontinuous Galerkin method for Navier Stokes is the so-called Symmetric Interior Penalty (SIP) method. It is used for example by Hartmann and Houston [21]. The symmetric nature of the discretization is particularly well suited with mesh adaptation by means of the adjoint equation resolution [22]. Last, we note that the discontinuous Galerkin method was already successfully tested in [16] at Direct Numerical Simulation scale for very moderate Reynolds, and also by Munz'team in Stuttgart [26], with local time stepping.

For concluding this section, there already exist numerical schemes based on the discontinuous Galerkin method which proved to be efficient for computing compressible viscous flows. Nevertheless, there remain things to be improved, which include: efficient shock capturing term methods for supersonic flows, high order discretization of curved boundaries, or low Mach behaviour of these schemes (this last point will be detailed in the next subsection). Another drawback of the discontinuous Galerkin methods is that they are very computationally costly, due to the accurate representation of the solution. A particular care must be taken on the implementation for being efficient.

3.3. Experimental aspects

A great deal of experiments has been devoted to the study of jet in crossflow configurations. They essentially differ one from each other by the hole shape (cylindrical or shaped), the hole axis inclination, the way by which the hole is fed, the characteristics of the crossflow and the jet (turbulent or not, isothermal or not), the number of holes considered and last but not least the techniques used to investigate the flow. A good starting point to assess the diversity of the studies carried out is given by [27]. For inclined cylindrical holes, the experimental

database produced by Gustafsson and Johansson [2] represents a sound reference base and for normal injection, the work by [34] served as reference for LES simulations [31]. For shaped holes, the studies are less numerous and are aimed at assessing the influence of the hole shape on various flow properties such as the heat transfer at the wall [24]. In 2007, Most [28] developed at UPPA a test facility for studying jet in crossflow issued from shaped holes . The hole shape was chosen as a 12.5 scale of the holes (i.e. at scale 1) drilled by laser in a combustion chamber. His preliminary 2-component PIV results have been used to test RANS simulations [29] and LES [30]. This test facility is extensively used in the framework of the present project to investigate a 1-hole jet i.e. an isolated jet in crossflow. PIV and LDV metrology are used.

[2] Slanted jet

CONCHA Project-Team

3. Research Program

3.1. Challenges related to numerical simulations of complex flows

First, we describe some typical difficulties in our fields of application which require the improvement of established and the development of new methods.

- Coupling of equations and models
  The general equations of fluid dynamics consist in a strongly coupled nonlinear system. Its mathematical nature depends on the precise model, but in general contains hyperbolic, parabolic, and elliptic parts. The spectrum of physical phenomena described by these equations is very large: convection, diffusion, waves... In addition, it is often necessary to couple different models in order to describe different parts of a mechanical system: chemistry, fluid-fluid-interaction, fluid-solid-interaction...
- Robustness with respect to physical parameters
  The values of physical parameters such as diffusion coefficients and constants describing different state equations and material laws lead to different behaviour characterized for example by the Reynolds, Mach, and Weissenberg numbers. Optimized numerical methods are available in many situations, but it remains a challenging problem in some fields of applications to develop robust discretizations and solution algorithms.
- Multiscale phenomena
  The inherent nonlinearities lead to an interplay of a wide range of physical modes, well-known for example from the study of turbulent flows. Since the resolution of all modes is often unreachable, it is a challenging task to develop numerical methods, which are still able to reproduce the essential features of the physical phenomenon under study.

3.2. Stabilized and discontinuous finite element methods

The discontinuous Galerkin method [61], [59], [40], [39] has gained enormous success in CFD due to its flexibility, links with finite volume methods, and its local conservation properties. In particular, it seems to be the most widely used finite element method for the Euler equations [41]. On the other hand, the main drawback of this approach is the large number of unknowns as compared to standard finite element methods. The situation is even worse if one counts the population of the resulting system matrices. In order to find a more efficient approach, it seems therefore important to study the connections with other finite element methods.

In view of the ubiquous problem of large Péclet numbers, stabilization techniques have been introduced since a long time. They are either based on upwinding or additional terms in the discrete variational formulation. The drawback of the first technique is a loss in consistency which generally leads to large numerical diffusion. The grand-father of the second technique is the SUPG/GLS method [50], [60]. Recently, new approaches have been developed, which try do avoid coupling of the different equations due to the residuals. In this context we cite LPS (local projection stabilization) [55], [49], [43][5] and CIP (continuous interior penalty) [51], [52].

3.3. Finite element methods on quadrilateral and hexahedral meshes

The construction of finite element methods on quadrilateral, and particularly, hexahedral meshes can be a complicated task; especially the development of mixed and non-conforming methods is an active field of research. The difficulties arise not only from the fact that adequate degrees of freedom have to be found, but also from the non-constantness of the element Jacobians; an arbitrary hexahedron, which we define as the image of the unit cube under a tri-linear transformation, does in general not have plane faces, which implies for example, that the normal vector is not constant on a side.

In collaboration with Eric Dubach (Associate professor at LMAP) and Jean-Marie Thomas (Former professor at LMAP) we have built a new class of finite element functions (named pseudo-conforming) on quadrilateral and hexahedral meshes. The degrees of freedom are the same as those of classical iso-parametric finite elements but the basis functions are defined as polynomials on each element of the mesh. On general quadrilaterals and hexahedra, our method leads to a non-conforming method; in the particular case of parallelotopes, the new finite elements coincide with the classical ones [54], [53].

3.4. Finite element methods for interface problems

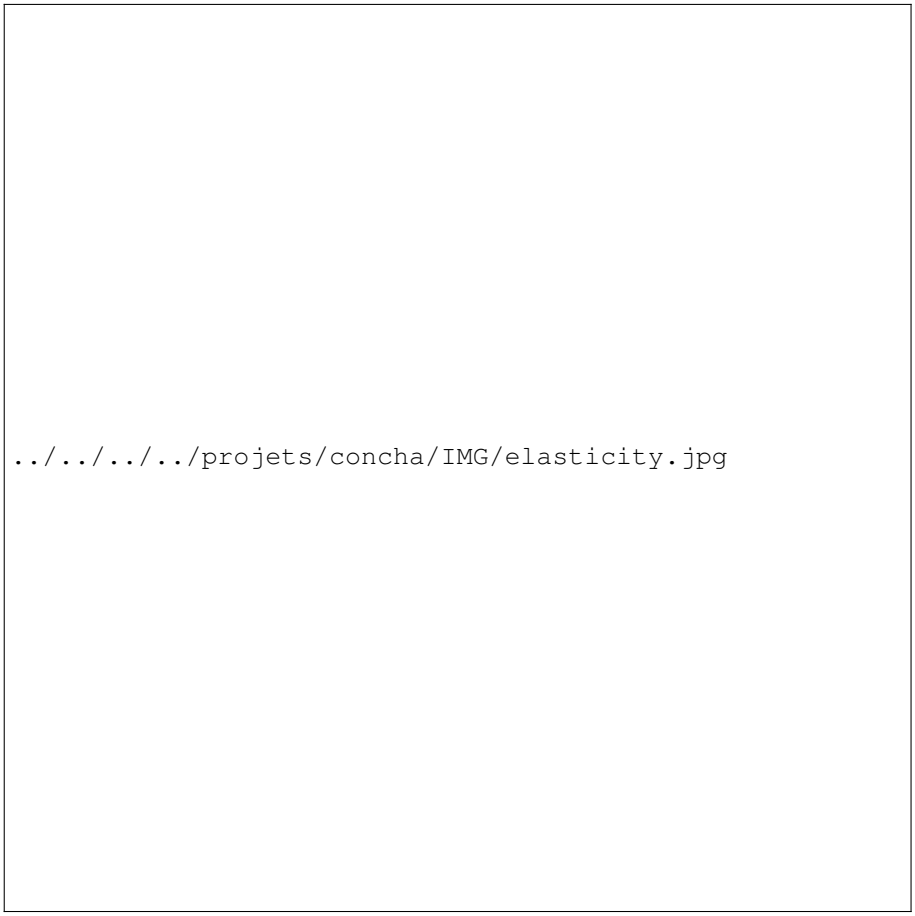


Figure 1. Incompressible elasticity with discontinuous material properties (left: modulus of velocities, right: pressure; from [42]).

The NXFEM (Nitsche eXtended finite element method) has been developed in [56] and [57]. It is based on a pure variational formulation with standard finite element spaces, which are locally enriched in such a way that the accurate capturing of an interface not aligned with the underlying mesh is possible, giving a rigorous formulation of the very popular XFEM. A typical computation for the Stokes problem with varying, piecewise constant viscosity is shown in Figure 1 . This technology opens the door to many applications in the field of fluid mechanics, such as immiscible flows, free surface flows and so on.

Figure 2. Solution with rough right-hand-side in a corner domain and adaptively refined mesh (from [45]).

3.5. Adaptivity

Adaptive finite element methods are becoming a standard tool in numerical simulations, and their application in CFD is one of the main topics of Concha. Such methods are based on a posteriori error estimates of the discretization error avoiding explicit knowledge of properties of the solution, in contrast to a priori error estimates. The estimator is used in an adaptive loop by means of a local mesh refinement algorithm. The mathematical theory of these algorithms has for a long time been bounded to the proof of upper and lower bounds, but has made important improvements in recent years. For illustration, a typical sequence of adaptively refined meshes on an L-shaped domain is shown in Figure 2 .

The theoretical analysis of mesh-adaptive methods, even in the most standard case of the Poisson problem, is in its infancy. The first important results in this direction concern the convergence of the sequence of solution generated by the algorithm (the standard a priori error analysis does not apply since the global mesh-size does not necessarily go to zero). In order to prove convergence, an unavoidable data approximation term has to be treated in addition to the error estimator [62]. These result do not say anything about the convergence speed, that is the number of unknowns required to achieve a given accuracy. Such complexity estimates are the subject of active research, the first fundamental result in this direction is [48].

Our first contribution [23] to this field has been the introduction of a new adaptive algorithm which makes use of an adaptive marking strategy, which refines according to the data oscillations only if they are by a certain factor larger then the estimator. This algorithm allowed us to prove geometric convergence and quasi-optimal complexity, avoiding additional iteration as used before [64]. We have extended our results to conforming FE without inner node refinement [46] and to mixed FE [45]. In this case, a major additional difficulty arises from the fact that, due to the saddle-point formulation, the orthogonality relation known from continuous FEM does not hold. In addition, we have considered the case of incomplete solution of the discrete systems. To this end, we have developed a simple adaptive stopping criterion based on comparison of the iteration error with the discretization error estimator, see also [44].

Goal-oriented error estimation has been introduced in [47]. It allows to error control and adaptivity directly oriented to the computation of physical quantities, such as the drag and lift coefficient, the Nusselt number, and other physical quantities.

3.6. LaTeX Test Page

Exemples d'équations :

- Equation en mode "mathématique" :
  $y = x^2$
- Equation en environnement "equation" :

$$P \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_r \end{pmatrix} = Q + R, \tag{1}$$

- Equation en environnement "displaymath" :

$$\sum_{0}^{\infty} y = x^4$$

- Autre exemple :

$$\forall f \in C^{\infty}\left(\left[-\frac{T}{2};\frac{T}{2}\right]\right), \forall t \in \left[-\frac{T}{2};\frac{T}{2}\right], f(\tau) = \sum_{k=-\infty}^{+\infty} e^{2i\pi\frac{k}{T}t} \times \underbrace{\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)e^{-2i\pi\frac{k}{T}t}dt}_{a_k = \widetilde{f}\left(\nu=\frac{k}{T}\right)}$$

Exemple de caractères spéciaux :
math pi : π
lettres : æ Æ à À â Â ä Ä ç Ç é É è È ê Ê ë Ë î Î ï Ï ô Ô ö Ö ù Ù û Û ü Ü ÿ þ Þ

Exemples d'images :

- Image en jpeg : voir image 3

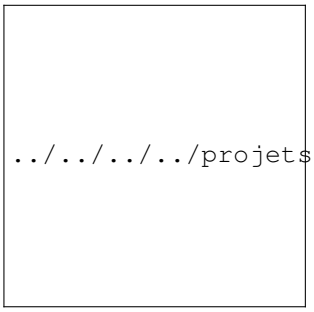


Figure 3. An example of a jpeg image

- Image en eps : voir figure 4

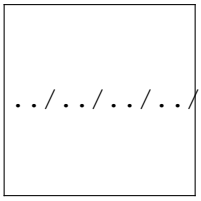


Figure 4. An example of an eps file

- Image en pdf : voir image 5

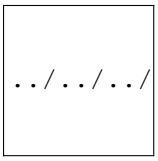


Figure 5. An example of a pdf file

CQFD Project-Team

3. Research Program

3.1. Introduction

The scientific objectives of the team are to provide mathematical tools for modeling and optimization of complex systems. These systems require mathematical representations which are in essence dynamic, multi-model and stochastic. This increasing complexity poses genuine scientific challenges in the domain of modeling and optimization. More precisely, our research activities are focused on stochastic optimization and (parametric, semi-parametric, multidimensional) statistics which are complementary and interlinked topics. It is essential to develop simultaneously statistical methods for the estimation and control methods for the optimization of the models.

3.2. Main research topics

- Stochastic modeling: Markov chain, Piecewise Deterministic Markov Processes (PDMP), Markov Decision Processes (MDP).
  The mathematical representation of complex systems is a preliminary step to our final goal corresponding to the optimization of its performance. For example, in order to optimize the predictive maintenance of a system, it is necessary to choose the adequate model for its representation. The step of modeling is crucial before any estimation or computation of quantities related to its optimization. For this we have to represent all the different regimes of the system and the behavior of the physical variables under each of these regimes. Moreover, we must also select the dynamic variables which have a potential effect on the physical variable and the quantities of interest. The team CQFD works on the theory of Piecewise Deterministic Markov Processes (PDMP's) and on Markov Decision Processes (MDP's). These two classes of systems form general families of controlled stochastic processes suitable for the modeling of sequential decision-making problems in the continuous-time (PDMPs) and discrete-time (MDP's) context. They appear in many fields such as engineering, computer science, economics, operations research and constitute powerful class of processes for the modeling of complex system.
- Estimation methods: estimation for PDMP; estimation in non- and semi parametric regression modeling.
  To the best of our knowledge, there does not exist any general theory for the problems of estimating parameters of PDMPs although there already exist a large number of tools for sub-classes of PDMPs such as point processes and marked point processes. However, to fill the gap between these specific models and the general class of PDMPs, new theoretical and mathematical developments will be on the agenda of the whole team. In the framework of non-parametric regression or quantile regression, we focus on kernel estimators or kernel local linear estimators for complete data or censored data. New strategies for estimating semi-parametric models via recursive estimation procedures have also received an increasing interest recently. The advantage of the recursive estimation approach is to take into account the successive arrivals of the information and to refine, step after step, the implemented estimation algorithms. These recursive methods do require restarting calculation of parameter estimation from scratch when new data are added to the base. The idea is to use only the previous estimations and the new data to refresh the estimation. The gain in time could be very interesting and there are many applications of such approaches.

- Dimension reduction: dimension-reduction via SIR and related methods, dimension-reduction via multidimensional and classification methods.
  Most of the dimension reduction approaches seek for lower dimensional subspaces minimizing the loss of some statistical information. This can be achieved in modeling framework or in exploratory data analysis context.
  In modeling framework we focus our attention on semi-parametric models in order to conjugate the advantages of parametric and nonparametric modeling. On the one hand, the parametric part of the model allows a suitable interpretation for the user. On the other hand, the functional part of the model offers a lot of flexibility. In this project, we are especially interested in the semi-parametric regression model $Y = f(X'\theta) + \varepsilon$, the unknown parameter θ belongs to $\mathbb{R}^p$ for a single index model, or is such that $\theta = [\theta_1, \cdots, \theta_d]$ (where each θ_k belongs to $\mathbb{R}^p$ and $d \leq p$ for a multiple indices model), the noise ε is a random error with unknown distribution, and the link function f is an unknown real valued function. Another way to see this model is the following: the variables X and Y are independent given $X'\theta$. In our semi-parametric framework, the main objectives are to estimate the parametric part θ as well as the nonparametric part which can be the link function f, the conditional distribution function of Y given X or the conditional quantile q_α. In order to estimate the dimension reduction parameter θ we focus on the Sliced Inverse Regression (SIR) method which has been introduced by Li [52] and Duan and Li [50]
  Methods of dimension reduction are also important tools in the field of data analysis, data mining and machine learning.They provide a way to understand and visualize the structure of complex data sets.Traditional methods among others are principal component analysis for quantitative variables or multiple component analysis for qualitative variables. New techniques have also been proposed to address these challenging tasks involving many irrelevant and redundant variables and often comparably few observation units. In this context, we focus on the problem of synthetic variables construction, whose goals include increasing the predictor performance and building more compact variables subsets. Clustering of variables is used for feature construction. The idea is to replace a group of "similar" variables by a cluster centroid, which becomes a feature. The most popular algorithms include K-means and hierarchical clustering. For a review, see, e.g., the textbook of Duda [51]
- Stochastic optimal control: optimal stopping, impulse control, continuous control, linear programming, singular perturbation, martingale problem.
  The first objective is to focus on the development of computational methods.
  - In the continuous-time context, stochastic control theory has from the numerical point of view, been mainly concerned with Stochastic Differential Equations (SDEs in short). From the practical and theoretical point of view, the numerical developments for this class of processes are extensive and largely complete. It capitalizes on the connection between SDEs and second order partial differential equations (PDEs in short) and the fact that the properties of the latter equations are very well understood. It is, however, hard to deny that the development of computational methods for the control of PDMPs has received little attention. One of the main reasons is that the role played by the familiar PDEs in the diffusion models is here played by certain systems of integro-differential equations for which there is not (and cannot be) a unified theory such as for PDEs as emphasized by M.H.A. Davis in his book. To the best knowledge of the team, there is only one attempt to tackle this difficult problem by O.L.V. Costa and M.H.A. Davis. The originality of our project consists in studying this unexplored area. It is very important to stress the fact that these numerical developments will give rise to a lot of theoretical issues such as type of approximations, convergence results, rates of convergence,....
  - Theory for MDP's has reached a rather high degree of maturity, although the classical tools such as value iteration, policy iteration and linear programming, and their various extensions, are not applicable in practice. We believe that the theoretical progress of MDP's must be in parallel with the corresponding numerical developments. Therefore, solving

MDP's numerically is an awkward and important problem both from the theoretical and practical point of view. In order to meet this challenge, the fields of neural networks, neuro-dynamic programming and approximate dynamic programming became recently an active area of research. Such methods found their roots in heuristic approaches, but theoretical results for convergence results are mainly obtained in the context of finite MDP's. Hence, an ambitious challenge is to investigate such numerical problems but for models with general state and action spaces. Our motivation is to develop theoretically consistent computational approaches for approximating optimal value functions and finding optimal policies.

Analysis of various problems arising in MDPs leads to a large variety of interesting mathematical problems. The second objective of the team is to study some theoretical aspects related to MDPs such as convex analytical methods and singular perturbation.

GEOSTAT Project-Team

3. Research Program

3.1. Dynamics of complex systems

GEOSTAT is studying complex signals under the point of view of *nonlinear* methods, in the sense of *nonlinear physics* i.e. the methodologies developed to study complex systems, with a strong emphasis on multiresolution analysis. Linear methods in signal processing refer to the standard point of view under which operators are expressed by simple convolutions with impulse responses. Linear methods in signal processing are widely used, from least-square deconvolution methods in adaptive optics to source-filter models in speech processing. Because of the absence of localization of the Fourier transform, linear methods are not successful to unlock the multiscale structures and cascading properties of variables which are of primary importance as stated by the physics of the phenomena. This is the reason why new approaches, such as DFA (Detrented Fluctuation Analysis), Time-frequency analysis, variations on curvelets [58] etc. have appeared during the last decades. Recent advances in dimensionality reduction, and notably in Compressive Sensing, go beyond the Nyquist rate in sampling theory using nonlinear reconstruction, but data reduction occur at random places, independently of geometric localization of information content, which can be very useful for acquisition purposes, but of lower impact in signal analysis. One important result obtained in GEOSTAT is the effective use of multiresolution analysis associated to optimal inference along the scales of a complex system. The multiresolution analysis is performed on dimensionless quantities given by the *singularity exponents* which encode properly the geometrical structures associated to multiscale organization. This is applied successfully in the derivation of high resolution ocean dynamics, or the high resolution mapping of gaseous exchanges between the ocean and the atmosphere; the latter is of primary importance for a quantitative evaluation of global warming. Understanding the dynamics of complex systems is recognized as a new discipline, which makes use of theoretical and methodological foundations coming from nonlinear physics, the study of dynamical systems and many aspects of computer science. One of the challenges is related to the question of *emergence* in complex systems: large-scale effects measurable macroscopically from a system made of huge numbers of interactive agents [50], [47], [63], [54]. Some quantities related to nonlinearity, such as Lyapunov exponents, Kolmogorov-Sinai entropy etc. can be computed at least in the phase space [48]. Consequently, knowledge from acquisitions of complex systems (which include *complex signals*) could be obtained from information about the phase space. A result from F. Takens [59] about strange attractors in turbulence has motivated the determination of discrete dynamical systems associated to time series [52], and consequently the theoretical determination of nonlinear characteristics associated to complex acquisitions. Emergence phenomena can also be traced inside complex signals themselves, by trying to localize information content geometrically. Fundamentally, in the nonlinear analysis of complex signals there are broadly two approaches: characterization by attractors (embedding and bifurcation) and time-frequency, multiscale/multiresolution approaches. Time-frequency analysis [49] and multiscale/multiresolution are the subjects of intense research and are profoundly reshaping the analysis of complex signals by nonlinear approaches [46], [51]. In real situations, the phase space associated to the acquisition of a complex phenomenon is unknown. It is however possible to relate, inside the signal's domain, local predictability to local reconstruction and deduce from that singularity exponents (SEs) [11] [7]. The SEs are defined at any point in the signal's domain, they relate, but are different, to other kinds of exponents used in the nonlinear analysis of complex signals. We are working on their relation with:

- properties in universality classses,
- the geometric localization of multiscale properties in complex signals,
- cascading characteristics of physical variables,
- optimal wavelets and inference in multiresolution analysis.

The alternative approach taken in GEOSTAT is microscopical, or geometrical: the multiscale structures which have their "fingerprint" in complex signals are being isolated in a single realization of the complex system, i.e. using the data of the signal itself, as opposed to the consideration of grand ensembles or a wide set of realizations. This is much harder than the ergodic approaches, but it is possible because a reconstruction formula such as the one derived in [60] is local and reconstruction in the signal's domain is related to predictability. This approach is analogous to the consideration of "microcanonical ensembles" in statistical mechanics.

Nonlinear signal processing is making use of quantities related to predictability. For instance the first Lyapunov exponent λ_1 is related, from Osedelec's theorem, to the limiting behaviour of the response, after a time t, to perturbation in the phase space $\log R_\tau(t)$:

$$\lambda_1 = \lim_{t\to\infty} \frac{1}{t} \langle \log R_\tau(t) \rangle \tag{2}$$

with $\langle \cdot \rangle$ being time average and R_τ the response to a perturbation [48]. In GEOSTAT our aim is to relate such classical quantities (among others) to the behaviour of SEs, which are defined by a limiting behaviour

$$\mu\left(\mathcal{B}_r(\mathbf{x})\right) \;=\; \alpha(\mathbf{x})\, r^{d+h(\mathbf{x})} \;+\; o\left(r^{d+h(\mathbf{x})}\right) \;\; (r \to 0) \tag{3}$$

(d: dimension of the signal's domain, μ: multiscale measure, typically whose density is the gradient's norm, $\mathcal{B}_r(\mathbf{x})$: ball of radius r centered at $\mathbf{x}$). For precise computation, SEs can be smoothly interpolated by projecting wavelets:

$$\mathcal{T}_\Psi \mu(\mathbf{x}, r) = \int_{\mathbb{R}^d} \mathrm{d}\mu(\mathbf{x}') \frac{1}{r^d} \Psi\left(\frac{\mathbf{x}-\mathbf{x}'}{r}\right) \tag{4}$$

(Ψ: mother wavelet, admissible or not), but the best numerical method in computing singularity exponents lies in the definition of a measure related to predictability [20], [43]:

$$h(\mathbf{x}) = \frac{\log \mathcal{T}_\Psi \mu(\mathbf{x}, r_0) / \langle \mathcal{T}_\Psi \mu(\cdot, r_0) \rangle}{\log r_0} + o(\frac{1}{\log r_0}) \tag{5}$$

with: r_0 is a scale choosen to diminish the amplitude of the correction term, and $\langle \mathcal{T}_\Psi \mu(\cdot, r_0) \rangle$ is the average value of the wavelet projection (mother wavelet Ψ) over the whole signal. Singularity exponents computed with this formula generalize the elementary "gradient's norm" in a very statistically coherent way across the scales.

SEs are related to the framework of reconstructible systems, and consequently to predictability. They unlock the geometric localization of a multiscale structure in a complex signal:

$$\mathcal{F}_h = \{\mathbf{x} \in \Omega \;\mid\; h(\mathbf{x}) = h\}, \tag{6}$$

(Ω: signal's domain). This multiscale structure is a fundamental feature of a complex system. Indeed, let us take the explicit example of a signal which is an acquisition of a 3D turbulent fluid. The velocity field of the flow, $\mathbf{v}(\mathbf{x},t)$, is a solution of the Navier-Stokes equations. Fully Developped Turbulence (FDT) is defined as the regime observed when the Reynolds number $R \to \infty$, R being defined as the ratio of "viscous diffusion time" by "circulation time": $R = \frac{LV}{\nu}$,L and V being respectively characteristic length and velocity of the flow. The phase space of the associated dynamical system is infinite dimensional, while the dynamics of the flow possess one or more finite dimensional attractors. In the case of FDT, particles of the fluid in the continuum which are trapped around KAM invariant manifolds undergo random perturbations in their motion which accounts for the "boost" observed in turbulent diffusion. From there comes the observed behaviour for the energy spectrum (the law $\mathcal{E}(\mathbf{k}) \sim |\mathbf{k}|^{-5/3}$ within the inertial range), an observation that was the starting point of the Kolmogorov K41 theory, but is still not directly mathematically related from the Navier-Stokes equations. Intermittency is observed within the inertial range and is related to the fact that, in the case of FDT, symmetry is restored only in a statistical sense, a fact that has consequences on the quality of any nonlinear signal representation by frames or dictionaries.

The example of FDT as a standard "template" for developing general methods that apply to a vast class of complex systems and signals is of fundamental interest because, in FDT, the existence of a multiscale hierarchy (i.e. the collection of sets $\mathcal{F}_h$ of equation 5) which is of multifractal nature and geometrically localized can be derived from physical considerations. This geometric hierarchy of sets is responsible for the shape of the computed singularity spectra, which in turn is related to the statistical organization of information content in a signal. It explains scale invariance, a characteristic feature of complex signals. The analogy from statistical physics comes from the fact that singularity exponents are direct generalizations of *critical exponents* which explain the macroscopic properties of a system around critical points, and the quantitative characterization of *universality classes*, which allow the definition of methods and algorithms that apply to general complex signals and systems, and not only turbulent signals: signals which belong to a same universality class share common statistical organization. In GEOSTAT, the approach to singularity exponents is done within a microcanonical setting, which can interestingly be compared with other approaches such that wavelet leaders, WTMM or DFA. During the past decades, classical approaches (here called "canonical" because they use the analogy taken from the consideration of "canonical ensembles" in statistical mechanics) permitted the development of a well-established analogy taken from thermodynamics in the analysis of complex signals: if $\mathcal{F}$ is the free energy, $\mathcal{T}$ the temperature measured in energy units, $\mathcal{U}$ the internal energy per volume unit $\mathcal{S}$ the entropy and $\widehat{\beta} = 1/\mathcal{T}$, then the scaling exponents associated to moments of intensive variables $p \to \tau_p$ corresponds to $\widehat{\beta}\mathcal{F}$, $\mathcal{U}(\widehat{\beta})$ corresponds to the singularity exponents values, and $\mathcal{S}(\mathcal{U})$ to the singularity spectrum.

The singularity exponents belong to a universality class, independently of microscopic properties in the phase space of various complex systems, and beyond the particular case of turbulent data (where the existence of a multiscale hierarchy, of multifractal nature, can be inferred directly from physical considerations). They describe common multiscale statistical organizations in different complex systems [57], and this is why GEOSTAT is working on nonlinear signal processing tools that are applied to very different types of signals. The methodological framework used in GEOSTAT for analyzing complex signals is different from, but related to, the "canonical" apparatus developed in recent years (WTMM method, wavelet leaders etc.). In the microcanonical approach developed, geometrically localized singularity exponents relate to a "microcanonical" description of multiplicative cascades observed in complex systems. Indeed, it can be shown that p-dissipation at scale r associated to a fixed interval $]p, p+\Delta p[$, $\epsilon_r^{(p,\Delta p)}$, behaves in the limit $\Delta p \to 0$ as

$$\epsilon_r^{(p)} = \lim_{\Delta p \to 0} \epsilon_r^{(p,\Delta p)} = \left(\epsilon_r^{(\infty)}\right)^{h(p)/h_\infty} \tag{7}$$

which indicates the existence of a relation between the multiscale hierarchy and the geometric localization of the cascade in complex systems.

The GEOSTAT team is working particularly on the very important subject of *optimal wavelets* which are wavelets ψ that "split" the signal projections between two different scales $\mathbf{r_1} < \mathbf{r_2}$ in such a way that there exists an injection term $\zeta_{r_1/r_2}(\mathbf{x})$, independent of the process $\mathcal{T}_\psi[\mathbf{s}](\mathbf{x},\mathbf{r})$ with:

$$\mathcal{T}_\psi[\mathbf{s}](\mathbf{x}, \mathbf{r_1}) = \zeta_{r_1/r_2}(\mathbf{x})\mathcal{T}_\psi[\mathbf{s}](\mathbf{x}, \mathbf{r_2}) \tag{8}$$

($\mathbf{r_1} < \mathbf{r_2}$: two scales of observation, ζ: injection variable between the scales, ψ: optimal wavelet). The **multiresolution analysis** associated to optimal wavelets is particularly interesting because it reflects, in an optimal way, the cross-scale information transfer in a complex system. These wavelets are related to persistence along the scales and lead to multiresolution analysis whose coefficients verify

$$\alpha_s = \eta_1 \alpha_f + \eta_2 \tag{9}$$

with α_s and α_f referring to child and parent coefficients, η_1 and η_2 are random variables independent of α_s and α_f and also independent of each other.

../../../../projets/geostat/IMG/image-rapp.png

Figure 1. Visualization of the motion field computed at high spatial resolution (pixel size: 4kms) over a wide area around South Africa. The ocean dynamics is computed by propagating low resolution information coming from altimetry data (pixel size: 24 kms) along approximated optimal multiresolution analysis computed over the singularity exponents of Sea Surface Temperature data obtained from MODIS AQUA and OSTIA. Common work between GEOSTAT and DYNBIO (LEGOS, CNRS UMR 55 66, Toulouse).

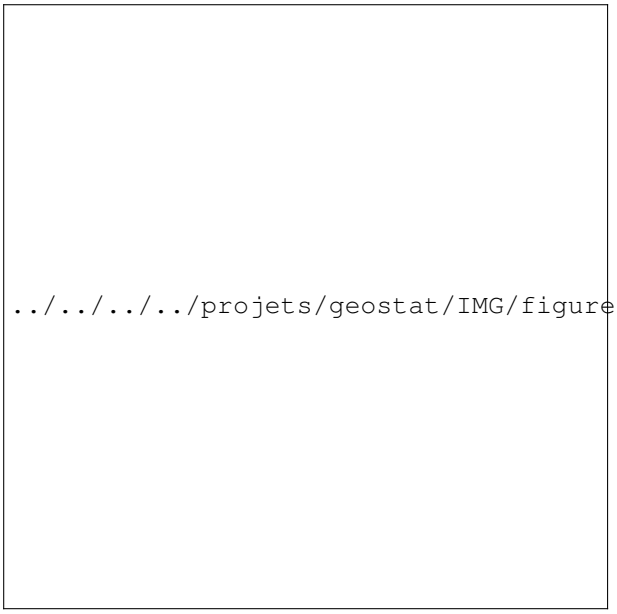
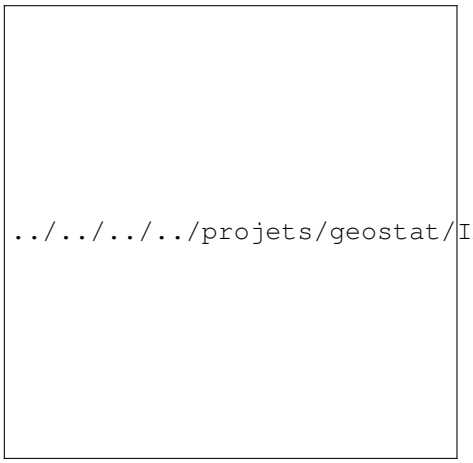


Figure 2. Result of the computation of a normalized source field over the 3D epicardial surface of the atria, from electric potential data acquired on a regular grid of electrodes placed on a patient's chest. There is a strong correlation between the red parts of the source field and the locations inside the heart where fibrillation occurs. Inputted data courtesy of IHU LIRYC.


../../../../projets/geostat/IMG/adaptive-optics1.png


../../../../projets/geostat/IMG/adaptive-optics3.png


../../../../projets/geostat/IMG/adaptive-optics2.png

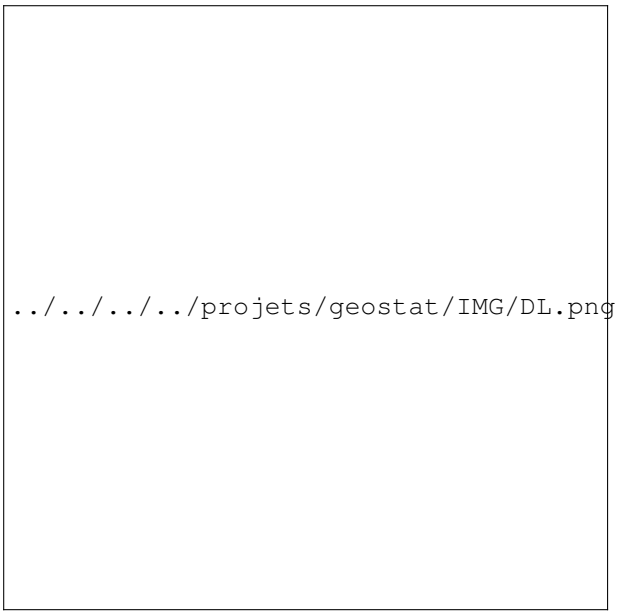


Figure 4. Top: A segment of a voiced speech signal (in black) along with the differentiated EGG (dEGG) recording (in red). Local maxima of dEGG shows the reference GCIs (yellow circles). Bottom: The singularity exponents (in blue) along with an auxiliary functional (in green) defined as $Zh(t) = \sum_{u=t-T_L}^{t-\delta t} h(u) - \sum_{u=t}^{t+T_L} h(u)$ *(*$h(t)$*: singularity exponent at* t*). In each positive half-period of* $Zh(t)$*, the minimum of singularity exponents is taken as the GCI (red circle).*

In a first example we give some insight about the collaboration with LEGOS Dynbio team [2] about high-resolution ocean dynamics from microcanonical formulations in nonlinear complex signal analysis. LPEs relate to the geometric structures linked with the cascading properties of indefinitely divisible variables in turbulent flows. Cascading properties can be represented by optimal wavelets (OWs); this opens new and fascinating directions of research for the determination of ocean motion field at high spatial resolution. OWs in a microcanonical sense pave the way for the determination of the energy injection mechanisms between the scales. From this results a new method for the complete evaluation of oceanic motion field; it consists in propagating along the scales the norm and the orientation of ocean dynamics deduced at low spatial resolution (geostrophic from altimetry and a part of ageostrophic from wind stress products). Using this approach, there is no need to use several temporal occurrences. Instead, the proper determination of the turbulent cascading and energy injection mechanisms in oceanographic signals allows the determination of oceanic motion field at the SST or Ocean colour spatial resolution (pixel size: 4 kms). We use the Regional Ocean Modelling System (ROMS) to validate the results on simulated data and compare the motion fields obtained with other techniques. See figure 1 .

In a second example, we show in figure 2 the highly promising results obtained in the application of nonlinear signal processing and multiscale techniques to the localization of heart fibrillation phenomenon acquired from a real patient and mapped over a reconstructed 3D surface of the heart. The notion of *source field*, defined in GEOSTAT from the computation of derivative measures related to the singularity exponents allows the localization of arrythmic phenomena inside the heart [8].

In a third example, we show in figure 3 the result of a new nonlinear method based on singularity exponents for optical phase reconstruction in adaptive optics (PhD of Suman Kumar Maji, defended November 2013). The method is very robust to noise. It consists in propagating subgradient information of acquired phase at low resolution across the scales of a multiresolution analysis computed on singularity exponents.

Our last example is about speech. In speech analysis, we use the concept of the Most Singular Manifold (MSM) to localize critical events in domain of this signal. We show that in case of voiced speech signals, the MSM coincides with the instants of significant excitation of the vocal tract system. It is known that these major excitations occur when the glottis is closed, and hence, they are called the Glottal Closure Instants (GCI). We use the MSM to develop a reliable and noise robust GCI detection algorithm and we evaluate our algorithm using contemporaneous Electro-Glotto-Graph (EGG) recordings. See figure 4 .

[2] http://www.legos.obs-mip.fr/recherches/equipes/dynbio.

MC2 Project-Team

3. Research Program

3.1. Introduction

We are mainly concerned with complex fluid mechanics problems. The complexity consists of the rheological nature of the fluids (non newtonian fluids), of the coupling phenomena (in shape optimization problems), of the geometry (micro-channels) or of multi-scale phenomena arising in turbulence or in tumor growth modeling. Our goal is to understand these phenomena and to simulate and/or to control them. The subject is wide and we will restrict ourselves to three directions: the first one consists in studying low Reynolds number interface problems in multi-fluid flows with applications to complex fluids, microfluidics and biology - the second one deals with numerical simulation of Newtonian fluid flows with emphasis on the coupling of methods to obtain fast solvers.

Even if we deal with several kinds of applications, there is a strong scientific core at each level of our project. Concerning the model, we are mainly concerned with incompressible flows and we work with the classical description of incompressible fluid dynamics. For the numerical methods, we use the penalization method to describe the obstacles or the boundary conditions for high Reynolds flows, for shape optimization, for interface problems in biology or in microfluidics. This allows us to use only cartesian meshes. Moreover, we use the level-set method for interface problems, for shape optimization and for fluid structure interaction. Finally, for the implementation, strong interaction exists between the members of the team and the modules of the numerical codes are used by all the team and we want to build the platform **eLYSe** to systematize this approach.

3.2. Multi-fluid flows and application for complex fluids, microfluidics

Participants: Angelo Iollo, Charles-Henri Bruneau, Thierry Colin, Mathieu Colin, Kévin Santugini.

Multi-fluid flows,microfluidics

By a complex fluid, we mean a fluid containing some mesoscopic objects, *i.e.* structures whose size is intermediate between the microscopic size and the macroscopic size of the experiment. The aim is to study complex fluids containing surfactants in large quantities. It modifies the viscosity properties of the fluids and surface-tension phenomena can become predominant.
Microfluidics is the study of fluids in very small quantities, in micro-channels (a micro-channel is typically 1 cm long with a section of $50\mu m \times 50\mu m$). They are many advantages of using such channels. First, one needs only a small quantity of liquid to analyze the phenomena. Furthermore, very stable flows and quite unusual regimes may be observed, which enables to perfom more accurate measurements. The idea is to couple numerical simulations with experiments to understand the phenomena, to predict the flows and compute some quantities like viscosity coefficients for example. Flows in micro-channels are often at low Reynolds numbers. The hydrodynamical part is therefore stable. However, the main problem is to produce real 3D simulations covering a large range of situations. For example we want to describe diphasic flows with surface tension and sometimes surface viscosity. Surface tension enforces the stability of the flow. The size of the channel implies that one can observe some very stable phenomena. For example, using a "T" junction, a very stable interface between two fluids can be observed. In a cross junction, one can also have formation of droplets that travel along the channel. Some numerical difficulties arise from the surface tension term. With an explicit discretization of this term, a restrictive stability condition appears for very slow flows [66]. Our partner is the LOF, a Rhodia-Bordeaux 1-CNRS laboratory.

One of the main points is the wetting phenomena at the boundary. Note that the boundary conditions are fundamental for the description of the flow since the channels are very shallow. The wetting properties cannot be neglected at all. Indeed, for the case of a two non-miscible fluids system, if one considers no-slip boundary conditions, then since the interface is driven by the velocity of the fluids, it shall not move on the boundary. The experiments shows that this is not true: the interface is moving and in fact all the dynamics start from the boundary and then propagate in the whole volume of fluids. Even with low Reynolds numbers, the wetting effects can induce instabilities and are responsible of hardly predictable flows. Moreover, the fluids that are used are often visco-elastic and exhibit "unusual" slip length. Therefore, we cannot use standard numerical codes and have to adapt the usual numerical methods to our case to take into account the specificities of our situations. In Johana Pinilla's thesis the Cox law has been implemented successfully to allow the interface to move properly between two Newtonian fluids of various viscosity or one Newtonian and one non-Newtonian fluid. Moreover, we want to obtain reliable models and simulations that can be as simple as possible and that can be used by our collaborators. As a summary, the main specific points of the physics are: the multi-fluid simulations at low Reynolds number, the wetting problems and the surface tension that are crucial, the 3D characteristic of the flows, the boundary conditions that are fundamental due to the size of the channels. We need to handle complex fluids. Our collaborators in this lab are H. Bordiguel, J.-B. Salmon, P. Guillot, A. Colin.

The evolution of non-newtonian flows in webs of micro-channels are therefore useful to understand the mixing of oil, water and polymer for enhanced oil recovery for example. Complex fluids arising in cosmetics are also of interest. We also need to handle mixing processes.

3.3. Cancer modeling

Participants: Sebastien Benzekry, Thierry Colin, Angelo Iollo, Clair Poignard, Olivier Saut, Lisl Weynans.

Tumor growth, cancer, metastasis

As in microfluidics, the growth of a tumor is a low Reynolds number flow. Several kinds of interfaces are present (membranes, several populations of cells,...) The biological nature of the tissues impose the use of different models in order to describe the evolution of tumor growth. The complexity of the geometry, of the rheological properties and the coupling with multi-scale phenomena is high but not far away from those encountered in microfluidics and the models and methods are close.

The challenge is twofold. On one hand, we wish to understand the complexity of the coupling effects between the different levels (cellular, genetic, organs, membranes, molecular). Trying to be exhaustive is of course hopeless, however it is possible numerically to isolate some parts of the evolution in order to better understand the interactions. Another strategy is to test *in silico* some therapeutic innovations. An example of such a test is given in [76] where the efficacy of radiotherapy is studied and in [77] where the effects of anti-invasive agents is investigated. It is therefore useful to model a tumor growth at several stage of evolution. The macroscopic continuous model is based on Darcy's law which seems to be a good approximation to describe the flow of the tumor cells in the extra-cellular matrix [45], [67], [68]. It is therefore possible to develop a two-dimensional model for the evolution of the cell densities. We formulate mathematically the evolution of the cell densities in the tissue as advection equations for a set of unknowns representing the density of cells with position (x, y) at time t in a given cycle phase. Assuming that all cells move with the same velocity given by Darcy's law and applying the principle of mass balance, one obtains the advection equations with a source term given by a cellular automaton. We assume diffusion for the oxygen and the diffusion constant depends on the density of the cells. The source of oxygen corresponds to the spatial location of blood vessels. The available quantities of oxygen interact with the proliferation rate given by the cellular automaton [76].

Another axis of our investigations in mathematical modeling-assisted theoretical biology is the biology and systemic dynamics of the metastatic process. This axis regroups several projects for which our approach can be decomposed into three steps. First, we base ourselves on a detailed study of the particular biological process, in close collaboration with biologists and the data they dispose. In a second step, we reduce the biological dynamics to its more essential components and build mathematical models able to simulate the process, to address the particular biological question under investigation and to give nontrivial insights on the overall

complex combination of these dynamics. Eventually, the last step consists in confronting the models to the data, using statistical parameter estimation methods, in order to identify theories or hypothesis that could or could not have generated the data and thus improve the biological understanding.

A forthcoming investigation in cancer treatment simulation is the influence of the electrochemotherapy [71] on the tumor growth. Electrochemotherapy consists in imposing to the malignant tumor high voltage electric pulses so that the plasma membrane of carcinoma cells is permeabilized. Biologically active molecules such as bleomycin, which usually cannot diffuse through the membrane, may then be internalized. A work in progress (C.Poignard [75] in collaboration with the CNRS lab of physical vectorology at the Institut Gustave Roussy) consists in modelling electromagnetic phenomena at the cell scale. A coupling between the microscopic description of the electroporation of cells and its influence on the global tumor growth at the macroscopic scale is expected. Another key point is the parametrization of the models in order to produce image-based simulations.

The second challenge is more ambitious. Mathematical models of cancer have been extensively developed with the aim of understanding and predicting tumor growth and the effects of treatments. In vivo modeling of tumors is limited by the amount of information available. However, in the last few years there have been dramatic increases in the range and quality of information available from non-invasive imaging methods, so that several potentially valuable imaging measurements are now available to quantitatively measure tumor growth, assess tumor status as well as anatomical or functional details. Using different methods such as the CT scan, magnetic resonance imaging (MRI), or positron emission tomography (PET), it is now possible to evaluate and define tumor status at different levels: physiological, molecular and cellular.

In this context, the present project aims at supporting the decision process of oncologists in the definition of therapeutic protocols via quantitative methods. The idea is to build mathematically and physically sound phenomenological models that can lead to patient-specific full-scale simulations, starting from data collected typically through medical imagery like CT scans, MRIs and PET scans or by quantitative molecular biology for leukemia. Our ambition is to provide medical doctors with patient-specific tumor growth models able to estimate, on the basis of previously collected data and within the limits of phenomenological models, the evolution at subsequent times of the pathology and possibly the response to the therapies.

The final goal is to provide numerical tools in order to help to answer to the crucial questions for a clinician:
When to start a treatment?
When to change a treatment?
When to stop a treatment?
Also we intend to incorporate real-time model information for improving the precision and effectiveness of non-invasive or micro-invasive tumor ablation techniques like acoustic hyperthermia, electroporation, radiofrequency or cryo-ablation.

We will specifically focus on the following pathologies: Lung and liver metastasis of a distant tumor
Low grade and high grade gliomas, meningiomas
Chronic myelogenous leukemia
These pathologies have been chosen because of the existing collaborations between the applied mathematics department of University of Bordeaux and the Institut Bergonié.

Our approach. Our approach is deterministic and spatial: it is based on solving an inverse problem based on imaging data. Models are of partial differential equation (PDE) type. They are coupled with a process of data assimilation based on imaging. We already have undertaken test cases on patients that are followed at Bergonié for lung metastases of thyroid tumors. These patients have a slowly evolving, asymptomatic metastatic disease, monitored by CT scans. On two thoracic images relative to successive times, the volume of the tumor under investigation is extracted by segmentation. To test our method, we chose patients without treatment and for whom we had at least three successive.

3.4. Newtonian fluid flows simulations and their analysis

Participants: Charles-Henri Bruneau, Angelo Iollo, Iraj Mortazavi, Michel Bergmann, Lisl Weynans.

Simulation, Analysis

It is very exciting to model complex phenomena for high Reynolds flows and to develop methods to compute the corresponding approximate solutions, however a well-understanding of the phenomena is necessary. Classical graphic tools give us the possibility to visualize some aspects of the solution at a given time and to even see in some way their evolution. Nevertheless in many situations it is not sufficient to understand the mechanisms that create such a behavior or to find the real properties of the flow. It is then necessary to carefully analyze the flow, for instance the vortex dynamics or to identify the coherent structures to better understand their impact on the whole flow behavior.
The various numerical methods used or developed to approximate the flows depend on the studied phenomenon. Our goal is to compute the most reliable method for each situation.
The first method, which is affordable in 2D, consists in a directly solving of the genuine Navier-Stokes equations in primitive variables (velocity-pressure) on Cartesian domains [54]. The bodies, around which the flow has to be computed are modeled using the penalization method (also named Brinkman-Navier-Stokes equations). This is an immersed boundary method in which the bodies are considered as porous media with a very small intrinsic permeability [46]. This method is very easy to handle as it consists only in adding a mass term U/K in the momentum equations. The boundary conditions imposed on artificial boundaries of the computational domains avoid any reflections when vortices cross the boundary. To make the approximation efficient enough in terms of CPU time, a multi-grid solver with a cell by cell Gauss-Seidel smoother is used.
The second type of methods is the vortex method. It is a Lagrangian technique that has been proposed as an alternative to more conventional grid-based methods. Its main feature is that the inertial nonlinear term in the flow equations is implicitly accounted for by the transport of particles. The method thus avoids to a large extent the classical stability/accuracy dilemma of finite-difference or finite-volume methods. This has been demonstrated in the context of computations for high Reynolds number laminar flows and for turbulent flows at moderate Reynolds numbers [61]. This method has recently enabled us to obtain new results concerning the three-dimensional dynamics of cylinder wakes.
The third method is to develop reduced order models (ROM) based on a Proper Orthogonal Decomposition (POD) [69]. The POD consists in approximating a given flow field $U(x,t)$ with the decomposition

$$U(x,t) = \sum_i a_i(t)\phi_i(x),$$

where the basis functions are empirical in the sense that they derive from an existing data base given for instance by one of the methods above. Then the approximation of Navier-Stokes equations for instance is reduced to solving a low-order dynamical system that is very cheap in terms of CPU time. Nevertheless the ROM can only restitute what is contained in the basis. Our challenge is to extend its application in order to make it an actual prediction tool.
The fourth method is a finite volume method on cartesian grids to simulate compressible Euler or Navier Stokes Flows in complex domains. An immersed boundary-like technique is developed to take into account boundary conditions around the obstacles with order two accuracy.

3.5. Flow control and shape optimization

Participants: Charles-Henri Bruneau, Angelo Iollo, Iraj Mortazavi, Michel Bergmann.

Flow Control, Shape Optimization

Flow simulations, optimal design and flow control have been developed these last years in order to solve real industrial problems : vortex trapping cavities with CIRA (Centro Italiano Ricerche Aerospaziali), reduction of vortex induced vibrations on deep sea riser pipes with IFP (Institut Français du Pétrole), drag reduction of a ground vehicle with Renault or in-flight icing with Bombardier and Pratt-Witney are some examples of possible applications of these researches. Presently the recent creation of the competitiveness cluster on aeronautics, space and embedded systems (AESE) based also in Aquitaine provides the ideal environment to extend our applied researches to the local industrial context. There are two main streams: the first need is to produce direct numerical simulations, the second one is to establish reliable optimization procedures.

In the next subsections we will detail the tools we will base our work on, they can be divided into three points: to find the appropriate devices or actions to control the flow; to determine an effective system identification technique based on the trace of the solution on the boundary; to apply shape optimization and system identification tools to the solution of inverse problems found in object imaging and turbomachinery.

3.5.1. Control of flows

There are mainly two approaches: passive (using passive devices on some specific parts that modify the shear forces) or active (adding locally some energy to change the flow) control.

The passive control consists mainly in adding geometrical devices to modify the flow. One idea is to put a porous material between some parts of an obstacle and the flow in order to modify the shear forces in the boundary layer. This approach may pose remarkable difficulties in terms of numerical simulation since it would be necessary, a priori, to solve two models: one for the fluid, one for the porous medium. However, by using the penalization method it becomes a feasible task [50]. This approach has been now used in several contexts and in particular in the frame of a collaboration with IFP to reduce vortex induced vibrations [51]. Another technique we are interested in is to inject minimal amounts of polymers into hydrodynamic flows in order to stabilize the mechanisms which enhance hydrodynamic drag.

The active approach is addressed to conceive, implement and test automatic flow control and optimization aiming mainly at two applications : the control of unsteadiness and the control and optimization of coupled systems. Implementation of such ideas relies on several tools. The common challenges are infinite dimensional systems, Dirichlet boundary control, nonlinear tracking control, nonlinear partial state observation.

The bottom-line to obtain industrially relevant control devices is the energy budget. The energy required by the actuators should be less than the energy savings resulting from the control application. In this sense our research team has gained a certain experience in testing several control strategies with a doctoral thesis (E. Creusé) devoted to increasing the lift on a dihedral plane. Indeed the extension of these techniques to real world problems may reveal itself very delicate and special care will be devoted to implement numerical methods which permit on-line computing of actual practical applications. For instance the method can be successful to reduce the drag forces around a ground vehicle and a coupling with passive control is under consideration to improve the efficiency of each control strategy.

3.5.2. System identification

We remark that the problem of deriving an accurate estimation of the velocity field in an unsteady complex flow, starting from a limited number of measurements, is of great importance in many engineering applications. For instance, in the design of a feedback control, a knowledge of the velocity field is a fundamental element in deciding the appropriate actuator reaction to different flow conditions. In other applications it may be necessary or advisable to monitor the flow conditions in regions of space which are difficult to access or where probes cannot be fitted without causing interference problems.

The idea is to exploit ideas similar to those at the basis of the Kalman filter. The starting point is again a Galerkin representation of the velocity field in terms of empirical eigenfunctions. For a given flow, the POD modes can be computed once and for all based on Direct Numerical Simulation (DNS) or on highly resolved experimental velocity fields, such as those obtained by particle image velocimetry. An instantaneous velocity field can thus be reconstructed by estimating the coefficients $a_i(t)$ of its Galerkin representation. One simple approach to estimate the POD coefficients is to approximate the flow measurements in a least square sense, as in [65].

A similar procedure is also used in the estimation based on gappy POD, see [80] and [84]. However, these approaches encounter difficulties in giving accurate estimations when three-dimensional flows with complicated unsteady patterns are considered, or when a very limited number of sensors is available. Under these conditions, for instance, the least squares approach cited above (LSQ) rapidly becomes ill-conditioned. This simply reflects the fact that more and more different flow configurations correspond to the same set of measurements.

Our challenge is to propose an approach that combines a linear estimation of the coefficients $a_i(t)$ with an appropriate non-linear low-dimensional flow model, that can be readily implemented for real time applications.

3.5.3. Shape optimization and system identification tools applied to inverse problems found in object imaging and turbomachinery

We will consider two different objectives. The first is strictly linked to the level set methods that are developed for microfluidics. The main idea is to combine different technologies that are developed with our team: penalization methods, level sets, an optimization method that regardless of the model equation will be able to solve inverse or optimization problems in 2D or 3D. For this we have started a project that is detailed in the research program. See also [57] for a preliminary application.

As for shape optimization in aeronautics, the aeroacoustic optimization problem of propeller blades is addressed by means of an inverse problem and its adjoint equations. This problem is divided into three subtasks:

i) formulation of an inverse problem for the design of propeller blades and determination of the design parameters ii) derivation of an aeroacoustic model able to predict noise levels once the blade geometry and the flow field are given iii) development of an optimization procedure in order to minimize the noise emission by controlling the design parameters.

The main challenge in this field is to move from simplified models [70] to actual 3D model. The spirit is to complete the design performed with a simplified tool with a fully three dimensional inverse problem where the load distribution as well as the geometry of the leading edge are those provided by the meridional plane analysis [79]. A 3D code will be based on the compressible Euler equations and an immersed boundary technique over a cartesian mesh. The code will be implicit and parallel, in the same spirit as what was done for the meridional plane. Further development include the extension of the 3D immersed boundary approach to time-dependent phenomena. This step will allow the designer to take into account noise sources that are typical of internal flows. The task will consist in including time dependent forcing on the inlet and/or outlet boundary under the form of Fourier modes and in computing the linearized response of the system. The optimization will then be based on a direct approach, i.e., an approach where the control is the geometry of the boundary. The computation of the gradient is performed by an adjoint method, which will be a simple "byproduct" of the implicit solver. The load distribution as well as the leading edge geometry obtained by the meridional plane approach will be considered as constraints of the optimization, by projection of the gradient on the constraint tangent plane. These challenges will be undertaken in collaboration with Politecnico di Torino and EC Lyon.

REALOPT Project-Team

3. Research Program

3.1. Introduction

Combinatorial optimization is the field of discrete optimization problems. In many applications, the most important decisions (control variables) are binary (on/off decisions) or integer (indivisible quantities). Extra variables can represent continuous adjustments or amounts. This results in models known as *mixed integer programs* (MIP), where the relationships between variables and input parameters are expressed as linear constraints and the goal is defined as a linear objective function. MIPs are notoriously difficult to solve: good quality estimations of the optimal value (bounds) are required to prune enumeration-based global-optimization algorithms whose complexity is exponential. In the standard approach to solving an MIP is so-called *branch-and-bound algorithm* : (i) one solves the linear programming (LP) relaxation using the simplex method; (ii) if the LP solution is not integer, one adds a disjunctive constraint on a factional component (rounding it up or down) that defines two sub-problems; (iii) one applies this procedure recursively, thus defining a binary enumeration tree that can be pruned by comparing the local LP bound to the best known integer solution. Commercial MIP solvers are essentially based on branch-and-bound (such IBM-CPLEX, FICO-Xpress-mp, or GUROBI). They have made tremendous progress over the last decade (with a speedup by a factor of 60). But extending their capabilities remains a continuous challenge; given the combinatorial explosion inherent to enumerative solution techniques, they remain quickly overwhelmed beyond a certain problem size or complexity.

Progress can be expected from the development of tighter formulations. Central to our field is the characterization of polyhedra defining or approximating the solution set and combinatorial algorithms to identify "efficiently" a minimum cost solution or separate an unfeasible point. With properly chosen formulations, exact optimization tools can be competitive with other methods (such as meta-heuristics) in constructing good approximate solutions within limited computational time, and of course has the important advantage of being able to provide a performance guarantee through the relaxation bounds. Decomposition techniques are implicitly leading to better problem formulation as well, while constraint propagation are tools from artificial intelligence to further improve formulation through intensive preprocessing. A new trend is robust optimization where recent progress have been made: the aim is to produce optimized solutions that remain of good quality even if the problem data has stochastic variations. In all cases, the study of specific models and challenging industrial applications is quite relevant because developments made into a specific context can become generic tools over time and see their way into commercial software.

Our project brings together researchers with expertise in mathematical programming (polyhedral approaches, Dantzig-Wolfe decomposition, mixed integer programing, robust and stochastic programming, and dynamic programming), graph theory (characterization of graph properties, combinatorial algorithms) and constraint programming in the aim of producing better quality formulations and developing new methods to exploit these formulations. These new results are then applied to find high quality solutions for practical combinatorial problems such as routing, network design, planning, scheduling, cutting and packing problems.

3.2. Polyhedral approaches for MIP

Adding valid inequalities to the polyhedral description of an MIP allows one to improve the resulting LP bound and hence to better prune the enumeration tree. In a cutting plane procedure, one attempt to identify valid inequalities that are violated by the LP solution of the current formulation and adds them to the formulation. This can be done at each node of the branch-and-bound tree giving rise to a so-called *branch-and-cut algorithm* [76]. The goal is to reduce the resolution of an integer program to that of a linear program by deriving a linear description of the convex hull of the feasible solutions. Polyhedral theory tells us that if X is a mixed integer program: $X = P \cap \mathbb{Z}^n \times \mathbb{R}^p$ where $P = \{x \in \mathbb{R}^{n+p} : Ax \leq b\}$ with matrix

$(A, b) \in \mathbb{Q}^{m \times (n+p+1)}$, then $conv(X)$ is a polyhedron that can be described in terms of linear constraints, i.e. it writes as $conv(X) = \{x \in \mathbb{R}^{n+p} : C\, x \leq d\}$ for some matrix $(C, d) \in \mathbb{Q}^{m' \times (n+p+1)}$ although the dimension m' is typically quite large. A fundamental result in this field is the equivalence of complexity between solving the combinatorial optimization problem $\min\{cx : x \in X\}$ and solving the *separation problem* over the associated polyhedron $conv(X)$: if $\widetilde{x} \notin conv(X)$, find a linear inequality $\pi\, x \geq \pi_0$ satisfied by all points in $conv(X)$ but violated by $\widetilde{x}$. Hence, for NP-hard problems, one can not hope to get a compact description of $conv(X)$ nor a polynomial time exact separation routine. Polyhedral studies focus on identifying some of the inequalities that are involved in the polyhedral description of $conv(X)$ and derive efficient *separation procedures* (cutting plane generation). Only a subset of the inequalities $C\, x \leq d$ can offer a good approximation, that combined with a branch-and-bound enumeration techniques permits to solve the problem. Using *cutting plane algorithm* at each node of the branch-and-bound tree, gives rise to the algorithm called *branch-and-cut*.

3.3. Decomposition and reformulation approaches

An hierarchical approach to tackle complex combinatorial problems consists in considering separately different substructures (subproblems). If one is able to implement relatively efficient optimization on the substructures, this can be exploited to reformulate the global problem as a selection of specific subproblem solutions that together form a global solution. If the subproblems correspond to subset of constraints in the MIP formulation, this leads to Dantzig-Wolfe decomposition. If it corresponds to isolating a subset of decision variables, this leads to Bender's decomposition. Both lead to extended formulations of the problem with either a huge number of variables or constraints. Dantzig-Wolfe approach requires specific algorithmic approaches to generate subproblem solutions and associated global decision variables dynamically in the course of the optimization. This procedure is known as *column generation*, while its combination with branch-and-bound enumeration is called *branch-and-price*. Alternatively, in Bender's approach, when dealing with exponentially many constraints in the reformulation, the *cutting plane procedures* that we defined in the previous section are well-suited tools. When optimization on a substructure is (relatively) easy, there often exists a tight reformulation of this substructure typically in an extended variable space. This gives rise powerful reformulation of the global problem, although it might be impractical given its size (typically pseudo-polynomial). It can be possible to project (part of) the extended formulation in a smaller dimensional space if not the original variable space to bring polyhedral insight (cuts derived through polyhedral studies can often be recovered through such projections).

3.4. Integration of Artificial Intelligence Techniques in Integer Programming

When one deals with combinatorial problems with a large number of integer variables, or tightly constrained problems, mixed integer programming (MIP) alone may not be able to find solutions in a reasonable amount of time. In this case, techniques from artificial intelligence can be used to improve these methods. In particular, we use primal heuristics and constraint programming.

Primal heuristics are useful to find feasible solutions in a small amount of time. We focus on heuristics that are either based on integer programming (rounding, diving, relaxation induced neighborhood search, feasibility pump), or that are used inside our exact methods (heuristics for separation or pricing subproblem, heuristic constraint propagation, ...).

Constraint Programming (CP) focuses on iteratively reducing the variable domains (sets of feasible values) by applying logical and problem-specific operators. The latter propagates on selected variables the restrictions that are implied by the other variable domains through the relations between variables that are defined by the constraints of the problem. Combined with enumeration, it gives rise to exact optimization algorithms. A CP approach is particularly effective for tightly constrained problems, feasibility problems and min-max problems Mixed Integer Programming (MIP), on the other hand, is known to be effective for loosely constrained problems and for problems with an objective function defined as the weighted sum of variables. Many problems belong to the intersection of these two classes. For such problems, it is reasonable to use algorithms that exploit complementary strengths of Constraint Programming and Mixed Integer Programming.

3.5. Robust Optimization

Decision makers are usually facing several sources of uncertainty, such as the variability in time or estimation errors. A simplistic way to handle these uncertainties is to overestimate the unknown parameters. However, this results in over-conservatism and a significant waste in resource consumption. A better approach is to account for the uncertainty directly into the decision aid model by considering mixed integer programs that involve uncertain parameters. Stochastic optimization account for the expected realization of random data and optimize an expected value representing the average situation. Robust optimization on the other hand entails protecting against the worst-case behavior of unknown data. There is an analogy to game theory where one considers an oblivious adversary choosing the realization that harms the solution the most. A full worst case protection against uncertainty is too conservative and induces very high over-cost. Instead, the realization of random data are bound to belong to a restricted feasibility set, the so-called uncertainty set. Stochastic and robust optimization rely on very large scale programs where probabilistic scenarios are enumerated. There is hope of a tractable solution for realistic size problems, provided one develops very efficient ad-hoc algorithms. The techniques for dynamically handling variables and constraints (column-and-row generation and Bender's projection tools) that are at the core of our team methodological work are specially well-suited to this context.

3.6. Polyhedral Combinatorics and Graph Theory

Many fundamental combinatorial optimization problems can be modeled as the search for a specific structure in a graph. For example, ensuring connectivity in a network amounts to building a *tree* that spans all the nodes. Inquiring about its resistance to failure amounts to searching for a minimum cardinality *cut* that partitions the graph. Selecting disjoint pairs of objects is represented by a so-called *matching*. Disjunctive choices can be modeled by edges in a so-called *conflict graph* where one searches for *stable sets* – a set of nodes that are not incident to one another. Polyhedral combinatorics is the study of combinatorial algorithms involving polyhedral considerations. Not only it leads to efficient algorithms, but also, conversely, efficient algorithms often imply polyhedral characterizations and related min-max relations. Developments of polyhedral properties of a fundamental problem will typically provide us with more interesting inequalities well suited for a branch-and-cut algorithm to more general problems. Furthermore, one can use the fundamental problems as new building bricks to decompose the more general problem at hand. For problem that let themselves easily be formulated in a graph setting, the graph theory and in particular graph decomposition theorem might help.

CARMEN Team

3. Research Program

3.1. Complex models for the propagation of cardiac action potentials

Cardiac arrythmias originates from the multiscale organisation of the cardiac action potential from the cellular scale up to the scale of the body. It relates the molecular processes from the cell membranes to the electrocardiogram, an electrical signal on the torso. The spatio-temporal patterns of this propagation is related both to the function of the cellular membrane and of the structural organisation of the cells into tissues, into the organ and final within the body.

Several improvements of current models of the propagation of the action potential will be developped, based on previous work [10], [2], [11] and on the data available at the LIRYC:

- Enrichment of the current monodomain and bidomain models by accounting for structural heterogeneities of the tissue at an intermediate scale. Here we focus on multiscale analysis techniques applied to the various high-resolution structural data available at the LIRYC.
- Coupling of the tissues from the different cardiac compartments and conduction systems. Here, we want to develop model that couples 1D, 2D and 3D phenomena described by reaction-diffusion PDEs.

These models are essential to improve our in-depth understanding of cardiac electrical dysfunction. To this aim, we will use high-performance computing techniques in order to explore numerically the complexity of these models and check that they are reliable experimental tools.

3.2. Simplified models and inverse problems

The medical and clinical exploration of the electrical signals is based on accurate reconstruction of the typical patterns of propagation of the action potential. The correct detection of these complex patterns by non-invasive electrical imaging techniques has to be developped. Both problems involve solving inverse problems that cannot be addressed with the more compex models. We want both to develop simple and fast models of the propagation of cardiac action potentials and improve the solutions to the inverse problems found in cardiac electrical imaging techniques.

The cardiac inverse problem consists in finding the cardiac activation maps or, more generally the whole cardiac electrical activity, from high density body surface electrocardiograms. It is a new and a powerful diagnosis technique, which success would be considered as a breakthrough in the cardiac diagnosis. Although widely studied during the last years, it remains a challenge for the scientific community. In many cases the quality of reconstructed electrical potential is not sufficiently accurate. The methods used consist in solving the Laplace equation on the volume delimited by the body surface and the epicardial surface.We plan to

- study in depth the dependance of this inverse problem inhomogeneities in the torso, conductivity values, the geometry, electrode placements...
- improve the solution to the inverse problem be using new regularization strategies and the theory of optimal control, both in the quasistatic and in the dynamic contexts.

Of cours we will use our models as a basis to regularize these inverse problems. We will conside the follwong strategies:

- using complete propagation models in the inverse problem, like the bidomain equations; for instance in order to localize some electrical sources;
- construct some families of reduced order models, using e.g. statistical learning techniques, which would accurately represent some families of well-identified pathologies;
- construct some simple models of the propagation of the activation front, based on eikonal or level-sets equations, but which would incorporate the representation of complex activation patterns.

Additionnaly, we will need to develop numerical techniques dedicated to our simplified eikonal/levl-sets equations.

3.3. Numerical techniques

We want the numerical simulations of the previous direct or inverse models to be efficient and reliable with respect to the need of the medical community. It needs to qualify and guarantee the accuracy and robustness of the numerical techniques and the efficiency of the resolution algorithms.

Based on previous work on solving the monodomain and bidomain equations [12], [13] and [15] and [1], we will focus on

- High-order numerical techniques with respect to the variables with physiological meaning, like velocity, AP duration and restitution properties;
- Efficient, dedicated preconditionning techniques coupled with parallel computing.

MAGIQUE-3D Project-Team

3. Research Program

3.1. Inverse Problems

- **Inverse scattering problems.** The determination of the shape of an obstacle immersed in a fluid medium from some measurements of the scattered field in the presence of incident waves is an important problem in many technologies such as sonar, radar, geophysical exploration, medical imaging and nondestructive testing. Because of its nonlinear and ill-posed character, this inverse obstacle problem (IOP) is very difficult to solve, especially from a numerical viewpoint. The success of the reconstruction depends strongly on the quantity and quality of the measurements, especially on the aperture (range of observation angles) and the level of noise in the data. Moreover, in order to solve IOP, the understanding of the theory for the associated direct scattering problem and the mastery of the corresponding solution methods are fundamental. Magique-3d is involved in the mathematical and numerical analysis of a direct elasto-acoustic scattering problem and of an inverse obstacle scattering problem. More specifically, the purpose of this research axis is to propose a solution methodology for the IOP based on a regularized Newton-type method, known to be robust and efficient.
- **Depth Imaging in the context of DIP.** The challenge of seismic imaging is to obtain an accurate representation of the subsurface from the solution of the full wave equation that is the best mathematical model according to the time reversibility of its solution. The Reverse Time Migration, [82], is a technique for Imaging which is widely used in the industry. It is an iterative process based on the solution of a collection of wave equations. The high complexity of the propagation medium requires the use of advanced numerical methods, which allows one to solve several wave equations quickly and accurately. Magique-3D is involved in Depth Imaging by the way of a collaboration with TOTAL, in the framework of the research program DIP which has been jointly defined by researchers of MAGIQUE-3D and engineers of TOTAL jointly. In this context, MAGIQUE-3D develops new algorithms in order to improve the RTM.

3.2. Modeling

The main activities of Magique-3D in modeling are the derivation and the analysis of models that are based on mathematical physics and are suggested by geophysical problems. In particular, Magique-3D considers equations of interest for the oil industry and focuses on the development and the analysis of numerical models which are well-adapted to solve quickly and accurately problems set in very large or unbounded domains as it is generally the case in geophysics.

- **Explicit High-Order Time Schemes.** Using the full wave equation for migration implies very high computational burdens, in order to get high resolution images. Indeed, to improve the accuracy of the numerical solution, one must considerably reduce the space step, which is the distance between two points of the mesh representing the computational domain. Another solution consists in using high-order finite element methods, which are very accurate even with coarse meshes. However, to take fully advantage of the high-order space discretization, one has to develop also high-order time schemes. The most popular ones for geophysical applications are the modified equation scheme [85], [100] and the ADER scheme [91]. Both rely on the same principle, which consists in applying a Taylor expansion in time to the solution of the wave equation. Then, the high-order derivatives with respect to the time are replaced by high order space operators, using the wave equation. Finally, auxiliary variables are introduced in order to transform the differential equation involving high-order operators into a system of differential equation with low order operators. The advantage of this technique is that it leads to explicit time schemes, which avoids the solution of huge linear

systems. The counterpart is that the schemes are only conditionally stable, which means that the time step is constrained by a CFL (Courant-Friedrichs-Levy) condition. The CFL number defines an upper bound for the time step in such a way that the smaller the space step is, the higher the numbers of iterations will be. Magique-3D is working on the construction and the analysis of new explicit time schemes which have either larger CFL numbers or local CFL numbers. By this way, the computational costs can be reduced without hampering the accuracy of the numerical solution.

- **Implicit High-Order Time Schemes.** Solving wave propagation problems in realistic media and in time domain is still a challenge. Implicit numerical schemes are nowadays considered as too expensive because they require the inversion of a linear system at each time step, contrary to the explicit schemes. However, explicit schemes are stable only when conditions on the discretization parameters are fulfilled, which can be very difficult to satisfy in realistic contexts and lead to very expensive simulations. These conditions become less dramatic or even disappear in some cases when using implicit schemes. Our goal is to construct, justify and optimize analytically original implicit schemes that seem accurate to solve specific difficulties coming from realistic problems. Several directions could be followed. First, we will continue to develop a methodology to construct high order implicit schemes for simple domains (conservative and homogeneous). For now (in [23]) we have used the modified equation technique on the classical θ-scheme, which leads to a parametrized family of numerical schemes that do not possess the same consistency error. Then, instead of choosing a time step that leads to a good precision for a given numerical scheme and spatial discretization, we reverse this standard reasoning and choose the best stable scheme, in the family of schemes that we just built, for a spatial and temporal given discretization. Stability is shown by energy techniques. It would be possible to continue this approach, leading to higher order schemes and better mastering the methodology. Crucial improvements to this work will be to adapt the methodology to dissipative media, heterogeneous media, realistic boundary conditions and model coupling. For instance, we aim at developing locally implicit schemes, for which the degree of implicitness would depend on the local characteristics of the media. Implicit-explicit schemes would be an application case of these new schemes, that could be used to optimize the global cost of simulation. Since computational efficiency is a priority, this theoretical seek will systematically be completed by the study of associated algorithms and their implementation on parallel architectures. We believe that locally implicit schemes will be well suited to the use of parallel algorithms.
- **Asymptotic methods for ultra short laser pulses propagation** In the long term goal of modeling an entire ultrashort laser chain, our first objective is to model the propagation of an ultrashort laser pulse in an isotropic third order nonlinear dispersive medium (as silica which is the material used for optical fibers or lenses). In other words, the optical index of the medium depends on the wavelength (dispersion phenomenon) but also on the electromagnetic field's intensity in a cubic way (Kerr effect). A first intuition is to use Maxwell's equations coupled with additional equations for the optical index. Current computing facilities allow us to solve such equations in parallel on small domains and during short time intervals, for instance using MONTJOIE software. The use of asymptotical methods that take advantage of the pulse's brevity leads to a family of equations written as evolution equations in the propagation direction (among which the nonlinear Schrödinger equation), and solved in frequency domain, which are much easier to solve. However, ultrashort pulses have large spectra, which contradicts another hypothesis currently done in usual asymptotic methods. This is why new models have to be derived, as well as numerical methods to solve them.

  In fiber optics, the laser pulse propagates inside a waveguide called "optical fiber", in which the transversal spatial repartition of the electromagnetic field can be shown to be a linear combination of eigenmodes. A first idea will be to generalize the results obtained in 1d (see 6.2.12) to this more realistic application. We have good reasons to believe that a very efficient model will be derived and will compare very well with the global Maxwell system. An ultimate validation will be obtained by comparing the numerical results with experimental data.

  Following this step, and in collaboration with CEA-CESTA, we wish to derive this kind of asymptotic models and associated numerical methods for general 3D open laser propagation.

- **Finite Element Methods for the time-harmonic wave equation.** As an alternative to Time-Domain Seismic Imaging, geophysicists are more and more interested by Time-Harmonic Seismic Imaging. The drawback of Time Domain Seismic Imaging is that it requires either to store the solution at each time step of the computation, or to perform many solutions to the wave equation. The advantage of Time Harmonic problems is that the solutions can be computed independently for each frequency and the images are produced with only two computations of the wave equation and without storing the solution. The counterpart is that one has to solve a huge linear system, which can not be achieved today when considering realistic 3D elastic media, even with the tremendous progress of Scientific Computing. Discontinuous Galerkin Methods (DGM), which are well-suited for hp-adaptivity, allow for the use of coarser meshes without hampering the accuracy of the solution. We are confident that these methods will help us to reduce the size of the linear system to be solved, but they still have to be improved in order to tackle realistic 3D problems. However, there exists many different DGMs, and the choice of the most appropriate one for geophysical applications is still not obvious. Our objectives are **a)** to propose a benchmark in order to test the performances of DGMs for seismic applications and **b)** to improve the most performant DGMs in order to be able to tackle realistic applications. To these aims, we propose to work in the following directions :
  1. To implement a 2D and 3D solver for time harmonic acoustic and elastodynamic wave equation, based on the Interior Penalty Discontinuous Galerkin Method(IPDGM). The implementation of this solver has started few years ago (see Section 5.1) for solving Inverse Scattering Problems and the results we obtained in 2D let us presage that IPDGM will be well-adapted for geophysical problems.
  2. To develop a new hybridizable DG (HDG) [84] for 2D and 3D elastodynamic equation. Instead of solving a linear system involving the degrees of freedom of all volumetric cells of the mesh, the principle of HDG consists in introducing a Lagrange multiplier representing the trace of the numerical solution on each face of the mesh. Hence, it reduces the number of unknowns of the global linear system and the volumetric solution is recovered thanks to a local computation on each element.
  3. To develop upscaling methods for very heterogeneous media. When the heterogenities are too small compared with the wavelengths of the waves, it is necessary to use such techniques, which are able to reproduce fine scale effects with computations on coarse meshes only.

  We also intend to consider finite elements methods where the basis functions are not polynomials, but solutions to the time-harmonic wave equations. We have already developed a numerical method based on plane wave basis functions [89]. The numerical results we have obtained on academic test cases showed that the proposed method is not only more stable than the DGM, but also exhibits a better level of accuracy. These results were obtained by choosing the same plane waves for the basis functions of every element of the mesh. We are now considering a new methodology allowing for the optimization of the angle of incidence of the plane waves at the element level.

  Last, we are developing an original numerical methods where the basis functions are fundamental solutions to the Helmholtz equation, such as Bessel or Hankel functions. Moreover, each basis function is not defined element by element but on the whole domain. This allows for reducing the volumetric variational formulation to a surfacic variational formulation.
- **Boundary conditions.** The construction of efficient absorbing boundary conditions (ABC) is very important for solving wave equations. Indeed, wave problems are generally set in unbounded or very large domains and simulation requires to limit the computational domain by introducing an external boundary, the so-called absorbing boundary. This topic has been a very active research topic during the past twenty years and despite that, efficient ABCs are have still to be designed. Classical conditions are constructed to absorb propagating waves and Magique-3D is investigating the way of improving existing ABCs by introducing the modelling of evanescent and glancing waves. For that purpose, we consider the micro-local derivation of the Dirichlet-to-Neumann operator. The interest of our approach is that the derivation does not depend on the geometry of the absorbing surface.

ABCs have been given up when Perfectly Matched Layers (PML) have been designed. PMLs have opened a large number of research directions and they are probably the most routinely used methods for modelling unbounded domains in geophysics. But in some cases, they turn out to be unstable. This is the case for some elastic media. We are thus considering the development of absorbing boundary conditions for elastodynamic media and in particular for Tilted Transverse Isotropic media, which are of high interest for geophysical applications.

- **Asymptotic modeling**.

  During the last 30 years, mathematicians have developed and justify approximate models with multiscale asymptotic analysis to deal with problems involving singularly perturbed geometry or problems with coefficients of different magnitude.

  Numerically, all these approximate models are of interest since they allow to mesh the computational domain without taking into account the small characteristic lengths. this techniques lead to a reduction of the computation burden. Unfortunately, these methods do not have penetrated the numerical community since most of the results have been obtained for the two dimensional Laplacian.

  The research activity of Magique 3D aims in extending this theory to three-dimensional challenging problems involving wave propagation phenomena. We address time harmonic and time dependent problems for acoustic waves, electromagnetic waves and elastodynamic wave which is a very important topic for industry. Moreover, it remains numerous open questions in the underlying mathematical problems.

  Another important issue is the modeling of boundary layers which are not governed by the same model than the rest of the computational domain. It is rather challenging to derive and to justify some matching condition between the boundary layer and the rest of the physical domain for such multiphysical problems.

  More precisely, we have worked in 2013 on the following topics:

  – Eddy current modeling in the context of electrothermic applications for the design of electromagnetic devices, in collaboration with laboratories Ampère, Laplace, Inria Team MC2, IRMAR, and F.R.S.-FNRS;
  – Multiphysic asymptotic modeling of multi perforate plates in turbo reactors in collaboration with Onera.
  – Modeling of small heterogeneities for the three dimensional time domain wave equation. This reduced models is a generalization of the so called Lax-Foldy reduced model.

3.3. High Performance methods for solving wave equations

Seismic Imaging of realistic 3D complex elastodynamic media does not only require advanced mathematical methods but also High Performing Computing (HPC) technologies, both from a software and hardware point of view. In the framework of our collaboration with Total, we are optimizing our algorithms, based on Discontinuous Galerkin methods, in the following directions.

- **Minimizing the communications between each processor.** One of the main advantages of Discontinuous Galerkin methods is that most of the calculi can be performed locally on each element of the mesh. The communications are carried out by the computations of fluxes on the faces of the elements. Hence, there are only communications between elements sharing a common face. This represents a considerable gain compared with Continuous Finite Element methods where the communications have to be done between elements sharing a common degree of freedom. However, the communications can still be minimized by judiciously choosing the quantities to be passed from one element to another.
- **Hybrid MPI and OpenMP parallel programming.** Since the communications are one of the main bottlenecks for the implementation of the Discontinuous Galerkin in an HPC framework, it is necessary to avoid these communications between two processors sharing the same RAM. To

this aim, the partition of the mesh is not performed at the core level but at the chip level and the parallelization between two cores of the same chip is done using OpenMP while the parallelization between two cores of two different chips is done using MPI.

- **Porting the code on new architectures.** We are now planning to port the code on the new Intel Many Integrated Core Architecture (Intel MIC). The optimization of this code has begun in 2013, in collaboration with Dider Rémy from SGI.
- **Using Runtimes Systems.** One of the main issue of optimization of parallel code is the portability between different architectures. Indeed, many optimizations performed for a specific architecture are often useless for another architecture. In some cases, they may even reduce the performance of the code. Task programming libraries such as StarPU (http://runtime.bordeaux.inria.fr/StarPU/ or DAGuE (http://icl.cs.utk.edu/dague/index.html) seem to be very promising to improve the portability of the code. These libraries handle the repartition of workloads between processors directly at the runtime level. However, until now, they have been mostly employed for solving linear algebra problems and we wish to test their performance on realistic wave propagation simulations. This is done in the framework of a collaboration with Inria Team Hiepacs and Georges Bosilca (University of Tennessee).

We are confident in the fact that the optimizations of the code will allow us to perform large-scale calculations and inversion of geophysical data for models and distributed data volumes with a resolution level impossible to reach in the past.

MAGNOME Project-Team

3. Research Program

3.1. Overview

Fundamental questions in the life sciences can now be addressed at an unprecedented scale through the combination of high-throughput experimental techniques and advanced computational methods from the computer sciences. The new field of *computational biology* or *bioinformatics* has grown around intense collaboration between biologists and computer scientists working towards understanding living organisms as *systems*. One of the key challenges in this study of systems biology is understanding how the static information recorded in the genome is interpreted to become dynamic systems of cooperating and competing biomolecules.

MAGNOME addresses this challenge through the development of informatic techniques for understanding the structure and history of eukaryote genomes: algorithms for genome analysis, data models for knowledge representation, stochastic hierarchical models for behavior of complex systems, and data mining and classification. Our work is in methods and algorithms for:

- **Genome annotation** for complete genomes, performing *syntactic* analyses to identify genes, and *semantic* analyses to map biological meaning to groups of genes [35], [6], [10], [11], [49], [50].
- **Integration of heterogenous data**, to build complete knowledge bases for storing and mining information from various sources, and for unambiguously exchanging this information between knowledge bases [1], [4], [41], [44], [33].
- **Ancestor reconstruction** using optimization techniques, to provide plausible scenarios of the history of genome evolution [11], [8], [45], [54].
- **Classification and logical inference**, to reliably identify similarities between groups of genetic elements, and infer rules through deduction and induction [9], [7], [10].
- **Hierarchical and comparative modeling**, to build mathematical models of the behavior of complex biological systems, in particular through combination, reutilization, and specialization of existing continuous and discrete models [40], [30], [53], [37], [52].

The hundred- to thousand-fold decrease in sequencing costs seen in the past few years presents significant challenges for data management and large-scale data mining. MAGNOME's methods specifically address "scaling out," where resources are added by installing additional computation nodes, rather than by adding more resources to existing hardware. Scaling out adds capacity and redundancy to the resource, and thus fault tolerance, by enforcing data redundancy between nodes, and by reassigning computations to existing nodes as needed.

3.2. Comparative genomics

The central dogma of evolutionary biology postulates that contemporary genomes evolved from a common ancestral genome, but the large scale study of their evolutionary relationships is frustrated by the unavailability of these ancestral organisms that have long disappeared. However, this common inheritance allows us to discover these relationships through *comparison*, to identify those traits that are common and those that are novel inventions since the divergence of different lineages.

We develop efficient methodologies and software for associating biological information with complete genome sequences, in the particular case where several phylogenetically-related eukaryote genomes are studied simultaneously.

The methods designed by MAGNOME for comparative genome annotation, structured genome comparison, and construction of integrated models are applied on a large scale to:

- eukaryotes from the hemiascomycete class of yeasts [49], [50], [6], [10], [2], [11] and to
- prokaryotes from the lactic bacteria used in winemaking [35], [36], [43], [34], [38], [32].

3.3. Comparative modeling

A general goal of systems biology is to acquire a detailed quantitative understanding of the dynamics of living systems. Different formalisms and simulation techniques are currently used to construct numerical representations of biological systems, and a recurring challenge is that hand-tuned, accurate models tend to be so focused in scope that it is difficult to repurpose them. We claim that, instead of modeling individual processes *de novo*, a sustainable effort in building efficient behavioral models must proceed incrementally. *Hierarchical modeling* is one way of combining specific models into networks. Effective use of hierarchical models requires both formal definition of the semantics of such composition, and efficient simulation tools for exploring the large space of complex behaviors. We have combined uses theoretical results from formal methods and practical considerations from modeling applications to define BioRica [27], [40], [53], a framework in which discrete and continuous models can communicate with a clear semantics. Hierarchical models in BioRica can be assembled from existing models, and translated into their execution semantics and then simulated at multiple resolutions through multi-scale stochastic simulation. BioRica models are compiled into a discrete event formalism capable of capturing discrete, continuous, stochastic, non deterministic and timed behaviors in an integrated and non-ambiguous way. Our long-term goal to develop a methodology in which we can **assemble a model** for a species of interest using a library of reusable models and a organism-level "schematic" determined by comparative genomics.

Comparative modeling is also a matter of reconciling experimental data with models [5] [30] and inferring new models through a combination of comparative genomics and successive refinement [46], [47].

MNEMOSYNE Team

3. Research Program

3.1. Integrative and Cognitive Neuroscience

The human brain is often considered as the most complex system dedicated to information processing. This multi-scale complexity, described from the metabolic to the network level, is particularly studied in integrative neuroscience, the goal of which is to explain how cognitive functions (ranging from sensorimotor coordination to executive functions) emerge from (are the result of the interaction of) distributed and adaptive computations of processing units, displayed along neural structures and information flows. Indeed, beyond the astounding complexity reported in physiological studies, integrative neuroscience aims at extracting, in simplifying models, regularities in space and functional mechanisms in time. From a spatial point of view, most neuronal structures (and particularly some of primary importance like the cortex, cerebellum, striatum, hippocampus) can be described through a regular organization of information flows and homogenous learning rules, whatever the nature of the processed information. From a temporal point of view, the arrangement in space of neuronal structures within the cerebral architecture also obeys a functional logic, the sketch of which is captured in models describing the main information flows in the brain, the corresponding loops built in interaction with the external and internal (bodily and hormonal) world and the developmental steps leading to the acquisition of elementary sensorimotor skills up to the most complex executive functions.

Three important characteristics are worth mentioning concerning these loops. Firstly, each of them sets a closed relation between the central nervous system and the rest of the world. This includes the external world (possibly including other intelligent agents), but also the internal world, with hormonal, physiological and bodily dimensions. Secondly, each of these loops can be described as a loop relating sensations to actions, in the wide sense of these terms: effectively, action can refer to acting in the real world, but also to modifying physiological parameters or controling neuronal activation. These loops have different constants of time, from immediate reflexes and sensorimotor adjustments to long term selection of motivation for action, the latter depending on hormonal and social parameters. Thirdly, each of the loops performs a learning reinforced by a primary (physiologically significant) or pseudo reward (sub-goal to be learned). As an illustration, we can mention respondent conditioning detecting stimuli anticipatory of primary rewards, episodic learning detecting multimodal events, and also more local phenomena like self-organization of topological structures. The gradual establishment of these loops and their mutual interactions give an interpretation of the resulting cognitive architecture as a synergetic system of memories.

In summary, integrative neuroscience builds, on an overwhelming quantity of data, a simplifying and interpretative grid suggesting homogenous local computations and a structured and logical plan for the development of cognitive functions. They arise from interactions and information exchange between neuronal structures and the external and internal world and also within the network of structures.

This domain is today very active and stimulating because it proposes, of course at the price of simplifications, global views of cerebral functioning and more local hypotheses on the role of subsets of neuronal structures in cognition. In the global approaches, the integration of data from experimental psychology and clinical studies leads to an overview of the brain as a set of interacting memories, each devoted to a specific kind of information processing [47]. It results also in longstanding and very ambitious studies for the design of cognitive architectures aiming at embracing the whole cognition. With the notable exception of works initiated by [43], most of these frameworks (e.g. Soar, ACT-R), though sometimes justified on biological grounds, do not go up to a *connectionist* neuronal implementation. Furthermore, because of the complexity of the resulting frameworks, they are restricted to simple symbolic interfaces with the internal and external world and to (relatively) small-sized internal structures. Our main research objective is undoubtly to build such a general purpose cognitive architecture (to model the brain *as a whole* in a systemic way), using a connectionist implementation and able to cope with a realistic environment.

3.2. Computational Neuroscience

From a general point of view, computational neuroscience can be defined as the development of methods from computer science and applied mathematics, to explore more technically and theoretically the relations between structures and functions in the brain [49], [36]. During the recent years this domain has gained an increasing interest in neuroscience and has become an essential tool for scientific developments in most fields in neuroscience, from the molecule to the system. In this view, all the objectives of our team can be described as possible progresses in computational neuroscience. Accordingly, it can be underlined that the systemic view that we promote can offer original contributions in the sense that, whereas most classical models in computational neuroscience focus on the better understanding of the structure/function relationship for isolated specific structures, we aim at exploring synergies between structures. Consequently, we target interfaces and interplay between heterogenous modes of computing, which is rarely addressed in classical computational neuroscience.

We also insist on another aspect of computational neuroscience which is, in our opinion, at the core of the involvement of computer scientists and mathematicians in the domain and on which we think we could particularly contribute. Indeed, we think that our primary abilities in numerical sciences imply that our developments are characterized above all by the effectiveness of the corresponding computations: We provide biologically inspired architectures with effective computational properties, such as robustness to noise, self-organization, on-line learning. We more generally underline the requirement that our models must also mimick biology through its most general law of homeostasis and self-adaptability in an unknown and changing environment. This means that we propose to numerically experiment such models and thus provide effective methods to falsify them.

Here, computational neuroscience means mimicking original computations made by the neuronal substratum and mastering their corresponding properties: computations are distributed and adaptive; they are performed without an homonculus or any central clock. Numerical schemes developed for distributed dynamical systems and algorithms elaborated for distributed computations are of central interest here [33], [42] and were the basis for several contributions in our group [48], [45], [50]. Ensuring such a rigor in the computations associated to our systemic and large scale approach is of central importance.

Equally important is the choice for the formalism of computation, extensively discussed in the connectionist domain. Spiking neurons are today widely recognized of central interest to study synchronization mechanisms and neuronal coupling at the microscopic level [34]; the associated formalism [39] can be possibly considered for local studies or for relating our results with this important domain in connectionism. Nevertheless, we remain mainly at the mesoscopic level of modeling, the level of the neuronal population, and consequently interested in the formalism developed for dynamic neural fields [31], that demonstrated a richness of behavior [35] adapted to the kind of phenomena we wish to manipulate at this level of description. Our group has a long experience in the study and adaptation of the properties of neural fields [45], [46] and their use for observing the emergence of typical cortical properties [38]. In the envisioned development of more complex architectures and interplay between structures, the exploration of mathematical properties such as stability and boundedness and the observation of emerging phenomena is one important objective. This objective is also associated with that of capitalizing our experience and promoting good practices in our software production (*cf.* § 5.1). In summary, we think that this systemic approach also brings to computational neuroscience new case studies where heterogenous and adaptive models with various time scales and parameters have to be considered jointly to obtain a mastered substratum of computation. This is particularly critical for large scale deployments, as we will discuss in § 5.1).

3.3. Machine Learning

The adaptive properties of the nervous system are certainly among its most fascinating characteristics, with a high impact on our cognitive functions. Accordingly, machine learning is a domain [41] that aims at giving such characteristics to artificial systems, using a mathematical framework (probabilities, statistics, data analysis, etc.). Some of its most famous algorithms are directly inspired from neuroscience, at different levels.

Connectionist learning algorithms implement, in various neuronal architectures, weight update rules, generally derived from the hebbian rule, performing non supervised (e.g. Kohonen self-organizing maps), supervised (e.g. layered perceptrons) or associative (e.g. Hopfield recurrent network) learning. Other algorithms, not necessarily connectionist, perform other kinds of learning, like reinforcement learning. Machine learning is a very mature domain today and all these algorithms have been extensively studied, at both the theoretical and practical levels, with much success. They have also been related to many functions (in the living and artificial domains) like discrimination, categorisation, sensorimotor coordination, planning, etc. and several neuronal structures have been proposed as the substratum for these kinds of learning [37], [30]. Nevertheless, we believe that, as for previous models, machine learning algorithms remain isolated tools, whereas our systemic approach can bring original views on these problems.

At the cognitive level, most of the problems we face do not rely on only one kind of learning and require instead skills that have to be learned in preliminary steps. That is the reason why cognitive architectures are often referred to as systems of memory, communicating and sharing information for problem solving. Instead of the classical view in machine learning of a flat architecture, a more complex network of modules must be considered here, as it is the case in the domain of deep learning. In addition, our systemic approach brings the question of incrementally building such a system, with a clear inspiration from developmental sciences. In this perspective, modules can generate internal signals corresponding to internal goals, predictions, error signals, able to supervise the learning of other modules (possibly endowed with a different learning rule), supposed to become autonomous after an instructing period. A typical example is that of episodic learning (in the hippocampus), storing declarative memory about a collection of past episods and supervising the training of a procedural memory in the cortex.

At the behavioral level, as mentionned above, our systemic approach underlines the fundamental links between the adaptive system and the internal and external world. The internal world includes proprioception and interoception, giving information about the body and its needs for integrity and other fundamental programs. The external world includes physical laws that have to be learned and possibly intelligent agents for more complex interactions. Both involve sensors and actuators that are the interfaces with these worlds and close the loops. Within this rich picture, machine learning generally selects one situation that defines useful sensors and actuators and a corpus with properly segmented data and time, and builds a specific architecture and its corresponding criteria to be satisfied. In our approach however, the first question to be raised is to discover what is the goal, where attention must be focused on and which previous skills must be exploited, with the help of a dynamic architecture and possibly other partners. In this domain, the behavioral and the developmental sciences, observing how and along which stages an agent learns, are of great help to bring some structure to this high dimensional problem.

At the implementation level, this analysis opens many fundamental challenges, hardly considered in machine learning : stability must be preserved despite on-line continuous learning; criteria to be satisfied often refer to behavioral and global measurements but they must be translated to control the local circuit level; in an incremental or developmental approach, how will the development of new functions preserve the integrity and stability of others? In addition, this continous re-arrangement is supposed to involve several kinds of learning, at different time scales (from msec to years in humans) and to interfer with other phenomena like variability and meta-plasticity.

In summary, our main objective in machine learning is to propose on-line learning systems, where several modes of learning have to collaborate and where the protocoles of training are realistic. We promote here a *really autonomous* learning, where the agent must select by itself internal resources (and build them if not available) to evolve at the best in an unknown world, without the help of any *deus-ex-machina* to define parameters, build corpus and define training sessions, as it is generally the case in machine learning. To that end, autonomous robotics (*cf.* § 3.4) is a perfect testbed.

3.4. Autonomous Robotics

Autonomous robots are not only convenient platforms to implement our algorithms; the choice of such platforms is also motivated by theories in cognitive science and neuroscience indicating that cognition emerges

from interactions of the body in direct loops with the world and develops interesting specificities accordingly. For example, internal representations can be minimized (opposite to building complex and hierarchical representations) and compensated by more simple strategies [32], more directly coupling perception and action and more efficient to react quickly in the changing environment (for example, instead of memorizing details of an object, just memorizing the eye movement to foveate it: the world itself is considered as an external memory). In this view for the *embodiment of cognition*, learning is intrinsically linked to sensorimotor loops and to a real body interacting with a real environment.

A real autonomy can be obtained only if the robot is able to define its goal by itself, without the specification of any high level and abstract cost function or rewarding state. To ensure such a capability, we propose to endow the robot with an artificial physiology, corresponding to perceive some kind of pain and pleasure. It may consequently discriminate internal and external goals (or situations to be avoided). This will mimick circuits related to fundamental needs (e.g. hunger and thirst) and to the preservation of bodily integrity. An important objective is to show that more abstract planning capabilities can arise from these basic goals.

A real autonomy with an on-line continuous learning as described in § 3.3 will be made possible by the elaboration of protocols of learning, as it is the case, in animal conditioning, for experimental studies where performance on a task can be obtained only after a shaping in increasingly complex tasks. Similarly, developmental sciences can teach us about the ordered elaboration of skills and their association in more complex schemes. An important challenge here is to translate these hints at the level of the cerebral architecture.

As a whole, autonomous robotics permits to assess the consistency of our models in realistic condition of use and offers to our colleagues in behavioral sciences an object of study and comparison, regarding behavioral dynamics emerging from interactions with the environment, also observable at the neuronal level.

In summary, our main contribution in autonomous robotics is to make autonomy possible, by various means corresponding to endow robots with an artificial physiology, to give instructions in a natural and incremental way and to prioritize the synergy between reactive and robust schemes over complex planning structures.

CEPAGE Project-Team

3. Research Program

3.1. Modeling Platform Dynamics

Modeling the platform dynamics in a satisfying manner, in order to design and analyze efficient algorithms, is a major challenge. In distributed platforms, the performance of individual nodes (be they computing or communication resources) will fluctuate; in a fully dynamic platform, the set of available nodes will also change over time, and algorithms must take these changes into account if they are to be efficient.

There are basically two ways one can model such evolution: one can use a *stochastic process*, or some kind of *adversary model*.

In a stochastic model, the platform evolution is governed by some specific probability distribution. One obvious advantage of such a model is that it can be simulated and, in many well-studied cases, analyzed in detail. The two main disadvantages are that it can be hard to determine how much of the resulting algorithm performance comes from the specifics of the evolution process, and that estimating how realistic a given model is – none of the current project participants are metrology experts.

In an adversary model, it is assumed that these unpredictable changes are under the control of an adversary whose goal is to interfere with the algorithms efficiency. Major assumptions on the system's behavior can be included in the form of restrictions on what this adversary can do (like maintaining such or such level of connectivity). Such models are typically more general than stochastic models, in that many stochastic models can be seen as a probabilistic specialization of a nondeterministic model (at least for bounded time intervals, and up to negligible probabilities of adopting "forbidden" behaviors).

Since we aim at proving guaranteed performance for our algorithms, we want to concentrate on suitably restricted adversary models. The main challenge in this direction is thus to describe sets of restricted behaviors that both capture realistic situations and make it possible to prove such guarantees.

3.2. Models for Platform Topology and Parameter Estimation

On the other hand, in order to establish complexity and approximation results, we also need to rely on a precise theoretical model of the targeted platforms.

- At a lower level, several models have been proposed to describe interference between several simultaneous communications. In the 1-port model, a node cannot simultaneously send to (and/or receive from) more than one node. Most of the "steady state" scheduling results have been obtained using this model. On the other hand, some authors propose to model incoming and outgoing communication from a node using fictitious incoming and outgoing links, whose bandwidths are fixed. The main advantage of this model, although it might be slightly less accurate, is that it does not require strong synchronization and that many scheduling problems can be expressed as multi-commodity flow problems, for which efficient decentralized algorithms are known. Another important issue is to model the bandwidth actually allocated to each communication when several communications compete for the same long-distance link.
- At a higher level, proving good approximation ratios on general graphs may be too difficult, and it has been observed that actual platforms often exhibit a simple structure. For instance, many real life networks satisfy small-world properties, and it has been proved, for instance, that greedy routing protocols on small world networks achieve good performance. It is therefore of interest to prove that logical (given by the interactions between hosts) and physical platforms (given by the network links) exhibit some structure in order to derive efficient algorithms.

3.3. General Framework for Validation

3.3.1. Low level modeling of communications

In the context of large scale dynamic platforms, it is unrealistic to determine precisely the actual topology and the contention of the underlying network at application level. Indeed, existing tools such as Alnem [114] are very much based on quasi-exhaustive determination of interferences, and it takes several days to determine the actual topology of a platform made up of a few tens of nodes. Given the dynamism of the platforms we target, we need to rely on less sophisticated models, whose parameters can be evaluated at runtime.

Therefore, we propose to model each node using a small set of parameters. This is related to the theoretical notion of distance labeling [103], and corresponds to assigning labels to the nodes, so that a cheap operation on the labels of two nodes provides an estimation of the value of a given parameter (the latency or the bandwidth between two nodes, for instance). Several solutions for performance estimation on the Internet are based on this notion, under the terminology of Network Coordinate Systems. Vivaldi [94], IDES [115] and Sequoia [117] are examples of such systems for latency estimation. In the case of bandwidth estimation, fewer solutions have been proposed. We have studied the last-mile model, in which we model each node by an incoming and an outgoing bandwidth and neglect interference that appears at the core of the network (Internet), in order to concentrate on local constraints.

3.3.2. Simulation

Once low level modeling has been obtained, it is crucial to be able to test the proposed algorithms. To do this, we will first rely on simulation rather than direct experimentation. Indeed, in order to be able to compare heuristics, it is necessary to execute those heuristics on the same platform. In particular, all changes in the topology or in the resource performance should occur at the same time during the execution of the different heuristics. In order to be able to replicate the same scenario several times, we need to rely on simulations. Moreover, a metric for providing approximation results in the case of dynamic platforms necessarily requires computing the optimal solution at each time step, which can be done off-line if all traces for the different resources are stored. Using simulation rather than experiments can be justified if the simulator itself has been proven valid. Moreover, the modeling of communications, processing and their interactions may be much more complex in the simulator than in the model used to provide a theoretical approximation ratio, such as in SimGrid. In particular, sophisticated TCP models for bandwidth sharing have been implemented in SimGRID.

During the course of the USS-SimGrid ANR Arpege project, the SimGrid simulation framework has been adapted to large scale environments. Thanks to hierarchical platform description, to simpler and more scalable network models, and to the possibility to distribute the simulation of several nodes, it is now possible to perform simulations of very large platforms (of the order of 10^5 resources). This work will be continued in the ANR SONGS project, which aims at making SimGrid usable for Next Generation Systems (P2P, Grids, Clouds, HPC). In this context, simulation of exascale systems are envisioned, and we plan to develop models for platform dynamicity to allow realistic and reproducible experimentation of our algorithms.

3.3.3. Practical validation and scaling

Finally, we propose several applications that will be described in detail in Section 5. These applications cover a large set of fields (molecular dynamics, continuous integration...). All these applications will be developed and tested with an academic or industrial partner. In all these collaborations, our goal is to prove that the services that we propose can be integrated as steering tools in already developed software. Our goal is to assert the practical interest of the services we develop and then to integrate and to distribute them as a library for large scale computing.

At a lower level, in order to validate the models we propose, i.e. make sure that the predictions given by the model are close enough to the actual values, we need realistic datasets of network performance on large scale distributed platforms. Latency measurements are easiest to perform, and several datasets are available to researchers and serve as benchmarks to the community. Bandwidth datasets are more difficult to obtain, because of the measurement cost. As part of the bedibe software (see section 5.4), we have implemented a

script to perform such measurements on the Planet-Lab platform [83]. We plan to make these datasets available to the community so that they can be used as benchmarks to compare the different solutions proposed.

HIEPACS Project-Team

3. Research Program

3.1. Introduction

The methodological component of HIEPACS concerns the expertise for the design as well as the efficient and scalable implementation of highly parallel numerical algorithms to perform frontier simulations. In order to address these computational challenges a hierarchical organization of the research is considered. In this bottom-up approach, we first consider in Section 3.2 generic topics concerning high performance computational science. The activities described in this section are transversal to the overall project and its outcome will support all the other research activities at various levels in order to ensure the parallel scalability of the algorithms. The aim of this activity is not to study general purpose solution but rather to address these problems in close relation with specialists of the field in order to adapt and tune advanced approaches in our algorithmic designs. The next activity, described in Section 3.3 , is related to the study of parallel linear algebra techniques that currently appear as promising approaches to tackle huge problems on extreme scale platforms. We highlight the linear problems (linear systems or eigenproblems) because they are in many large scale applications the main computational intensive numerical kernels and often the main performance bottleneck. These parallel numerical techniques, which are involved in the IPL C2S@EXA, will be the basis of both academic and industrial collaborations, some are described in Section 4.1 , but will also be closely related to some functionalities developed in the parallel fast multipole activity described in Section 3.4 . Finally, as the accuracy of the physical models increases, there is a real need to go for parallel efficient algorithm implementation for multiphysics and multiscale modeling in particular in the context of code coupling. The challenges associated with this activity will be addressed in the framework of the activity described in Section 3.5 .

Currently, we have one major application (see Section 4.1) that is in material physics. We will contribute to all steps of the design of the parallel simulation tool. More precisely, our applied mathematics skill will contribute to the modelling, our advanced numerical schemes will help in the design and efficient software implementation for very large parallel multi-scale simulations. We also participate to a few co-design actions in close collaboration with some applicative groups. The objective of this activity is to instanciate our expertise in fields where they are critical for designing scalable simulation tools. We refer to Section 4.2 for a detailed description of these activities.

3.2. High-performance computing on next generation architectures

Participants: Emmanuel Agullo, Olivier Coulaud, Luc Giraud, Mathieu Faverge, Abdou Guermouche, Matías Hastaran, Andra Hugo, Xavier Lacoste, Guillaume Latu, Stojce Nakov, Florent Pruvost, Pierre Ramet, Jean Roman, Mawussi Zounon.

The research directions proposed in HIEPACS are strongly influenced by both the applications we are studying and the architectures that we target (i.e., massively parallel many-core architectures, ...). Our main goal is to study the methodology needed to efficiently exploit the new generation of high-performance computers with all the constraints that it induces. To achieve this high-performance with complex applications we have to study both algorithmic problems and the impact of the architectures on the algorithm design.

From the application point of view, the project will be interested in multiresolution, multiscale and hierarchical approaches which lead to multi-level parallelism schemes. This hierarchical parallelism approach is necessary to achieve good performance and high-scalability on modern massively parallel platforms. In this context, more specific algorithmic problems are very important to obtain high performance. Indeed, the kind of applications we are interested in are often based on data redistribution for example (e.g. code coupling applications). This well-known issue becomes very challenging with the increase of both the number of computational nodes and the amount of data. Thus, we have both to study new algorithms and to adapt the

existing ones. In addition, some issues like task scheduling have to be restudied in this new context. It is important to note that the work done in this area will be applied for example in the context of code coupling (see Section 3.5).

Considering the complexity of modern architectures like massively parallel architectures or new generation heterogeneous multicore architectures, task scheduling becomes a challenging problem which is central to obtain a high efficiency. Of course, this work requires the use/design of scheduling algorithms and models specifically to tackle our target problems. This has to be done in collaboration with our colleagues from the scheduling community like for example O. Beaumont (Inria REALOPT Project-Team). It is important to note that this topic is strongly linked to the underlying programming model. Indeed, considering multicore architectures, it has appeared, in the last five years, that the best programming model is an approach mixing multi-threading within computational nodes and message passing between them. In the last five years, a lot of work has been developed in the high-performance computing community to understand what is critic to efficiently exploit massively multicore platforms that will appear in the near future. It appeared that the key for the performance is firstly the grain of computations. Indeed, in such platforms the grain of the parallelism must be small so that we can feed all the processors with a sufficient amount of work. It is thus very crucial for us to design new high performance tools for scientific computing in this new context. This will be developed in the context of our solvers, for example, to adapt to this new parallel scheme. Secondly, the larger the number of cores inside a node, the more complex the memory hierarchy. This remark impacts the behaviour of the algorithms within the node. Indeed, on this kind of platforms, NUMA effects will be more and more problematic. Thus, it is very important to study and design data-aware algorithms which take into account the affinity between computational threads and the data they access. This is particularly important in the context of our high-performance tools. Note that this work has to be based on an intelligent cooperative underlying run-time (like the tools developed by the Inria RUNTIME Project-Team) which allows a fine management of data distribution within a node.

Another very important issue concerns high-performance computing using “heterogeneous” resources within a computational node. Indeed, with the emergence of the GPU and the use of more specific co-processors, it is important for our algorithms to efficiently exploit these new kind of architectures. To adapt our algorithms and tools to these accelerators, we need to identify what can be done on the GPU for example and what cannot. Note that recent results in the field have shown the interest of using both regular cores and GPU to perform computations. Note also that in opposition to the case of the parallelism granularity needed by regular multicore architectures, GPU requires coarser grain parallelism. Thus, making both GPU and regular cores work all together will lead to two types of tasks in terms of granularity. This represents a challenging problem especially in terms of scheduling. From this perspective, we investigate new approaches for composing parallel applications within a runtime system for heterogeneous platforms.

The SOLHAR project aims at studying and designing algorithms and parallel programming models for implementing direct methods for the solution of sparse linear systems on emerging computers equipped with accelerators. Several attempts have been made to accomplish the porting of these methods on such architectures; the proposed approaches are mostly based on a simple offloading of some computational tasks (the coarsest grained ones) to the accelerators and rely on fine hand-tuning of the code and accurate performance modeling to achieve efficiency. SOLHAR proposes an innovative approach which relies on the efficiency and portability of runtime systems, such as the StarPU tool developed in the RUNTIME team. Although the SOLHAR project will focus on heterogeneous computers equipped with GPUs due to their wide availability and affordable cost, the research accomplished on algorithms, methods and programming models will be readily applicable to other accelerator devices. Our final goal would be to have high performance solvers and tools which can efficiently run on all these types of complex architectures by exploiting all the resources of the platform (even if they are heterogeneous).

In order to achieve an advanced knowledge concerning the design of efficient computational kernels to be used on our high performance algorithms and codes, we will develop research activities first on regular frameworks before extending them to more irregular and complex situations. In particular, we will work first on optimized dense linear algebra kernels and we will use them in our more complicated direct and hybrid

solvers for sparse linear algebra and in our fast multipole algorithms for interaction computations. In this context, we will participate to the development of those kernels in collaboration with groups specialized in dense linear algebra. In particular, we intend develop a strong collaboration with the group of Jack Dongarra at the University of Tennessee and collaborating research groups. The objectives will be to develop dense linear algebra algorithms and libraries for multicore architectures in the context the PLASMA project and for GPU and hybrid multicore/GPU architectures in the context of the MAGMA project. The framework that hosts all these research activities is the associated team MORSE.

A more prospective objective is to study the fault tolerance in the context of large-scale scientific applications for massively parallel architectures. Indeed, with the increase of the number of computational cores per node, the probability of a hardware crash on a core is dramatically increased. This represents a crucial problem that needs to be addressed. However, we will only study it at the algorithmic/application level even if it needed lower-level mechanisms (at OS level or even hardware level). Of course, this work can be done at lower levels (at operating system) level for example but we do believe that handling faults at the application level provides more knowledge about what has to be done (at application level we know what is critical and what is not). The approach that we will follow will be based on the use of a combination of fault-tolerant implementations of the run-time environments we use (like for example FT-MPI) and an adaptation of our algorithms to try to manage this kind of faults. This topic represents a very long range objective which needs to be addressed to guaranty the robustness of our solvers and applications. In that respect, we are involved in a ANR-Blanc project entitles RESCUE jointly with two other Inria EPI, namely ROMA and GRAND-LARGE and the G8 ESC international initiative. The main objective of the RESCUE project is to develop new algorithmic techniques and software tools to solve the exascale resilience problem. Solving this problem implies a departure from current approaches, and calls for yet-to-be- discovered algorithms, protocols and software tools.

Finally, it is important to note that the main goal of HIEPACS is to design tools and algorithms that will be used within complex simulation frameworks on next-generation parallel machines. Thus, we intend with our partners to use the proposed approach in complex scientific codes and to validate them within very large scale simulations.

3.3. High performance solvers for large linear algebra problems

Participants: Emmanuel Agullo, Astrid Casadei, Olivier Coulaud, Mathieu Faverge, Romain Garnier, Luc Giraud, Abdou Guermouche, Andra Hugo, Pablo Salas Medina, Stojce Nakov, Julien Pedron, Florent Pruvost, Pierre Ramet, Jean Roman, Xavier Vasseur.

Starting with the developments of basic linear algebra kernels tuned for various classes of computers, a significant knowledge on the basic concepts for implementations on high-performance scientific computers has been accumulated. Further knowledge has been acquired through the design of more sophisticated linear algebra algorithms fully exploiting those basic intensive computational kernels. In that context, we still look at the development of new computing platforms and their associated programming tools. This enables us to identify the possible bottlenecks of new computer architectures (memory path, various level of caches, inter processor or node network) and to propose ways to overcome them in algorithmic design. With the goal of designing efficient scalable linear algebra solvers for large scale applications, various tracks will be followed in order to investigate different complementary approaches. Sparse direct solvers have been for years the methods of choice for solving linear systems of equations, it is nowadays admitted that classical approaches are not scalable neither from a computational complexity nor from a memory view point for large problems such as those arising from the discretization of large 3D PDE problems. We will continue to work on sparse direct solvers on one hand to make sure they fully benefit from most advanced computing platforms on the other hand because they are a key building boxes for the design of some of our parallel algorithms such as the hybrid solvers described in the sequel of this section. Our activities in that context will mainly address preconditioned Krylov subspace methods; both components, preconditioner and Krylov solvers, will be investigated. In this framework, and possibly in relation with the research activity on fast multipole, we intend to study how emerging H-matrix arithmetic can benefit to our solver research efforts.

3.3.1. Parallel sparse direct solver

Solving large sparse systems $Ax = b$ of linear equations is a crucial and time-consuming step, arising in many scientific and engineering applications. Consequently, many parallel techniques for sparse matrix factorization have been studied and implemented.

Sparse direct solvers are mandatory when the linear system is very ill-conditioned; such a situation is often encountered in structural mechanics codes, for example. Therefore, to obtain an industrial software tool that must be robust and versatile, high-performance sparse direct solvers are mandatory, and parallelism is then necessary for reasons of memory capability and acceptable solution time. Moreover, in order to solve efficiently 3D problems with more than 50 million unknowns, which is now a reachable challenge with new multicore supercomputers, we must achieve good scalability in time and control memory overhead. Solving a sparse linear system by a direct method is generally a highly irregular problem that induces some challenging algorithmic problems and requires a sophisticated implementation scheme in order to fully exploit the capabilities of modern supercomputers.

New supercomputers incorporate many microprocessors which include themselves one or many computational cores. These new architectures induce strongly hierarchical topologies. These are called NUMA architectures. In the context of distributed NUMA architectures, in collaboration with the Inria RUNTIME team, we study optimization strategies to improve the scheduling of communications, threads and I/O. We have developed dynamic scheduling designed for NUMA architectures in the PaStiX solver. The data structures of the solver, as well as the patterns of communication have been modified to meet the needs of these architectures and dynamic scheduling. We are also interested in the dynamic adaptation of the computation grain to use efficiently multi-core architectures and shared memory. Experiments on several numerical test cases have been performed to prove the efficiency of the approach on different architectures.

In collaboration with the ICL team from the University of Tennessee, and the RUNTIME team from Inria, we are evaluating the way to replace the embedded scheduling driver of the PaStiX solver by one of the generic frameworks, PaRSEC or StarPU, to execute the task graph corresponding to a sparse factorization. The aims is to design algorithms and parallel programming models for implementing direct methods for the solution of sparse linear systems on emerging computer equipped with GPU accelerators. More generally, this work will be performed in the context of the ANR SOLHAR project which aims at designing high performance sparse direct solvers for modern heterogeneous systems. The project involves several groups working either on the sparse linear solver aspects (HIEPACS and ROMA from Inria and APO from IRIT), on runtime systems (RUNTIME from Inria) or scheduling algorithms (REALOPT and ROMA from Inria). The results of these efforts will be validated in the applications provided by the industrial project members, namely CEA-CESTA and EADS-IW.

On the numerical side, we are studying how the data sparsness that might exist in some dense blocks appearing during the factorization can be exploited using different compression techniques based on H-matrix (and variants) arithmetics. This research activity will be conducted in the framework of the FASTLA associate team and will naturally irrigate the hybrid solvers described below as well as closely interact with the other research efforts where similar data sparsness might be exploited.

3.3.2. Hybrid direct/iterative solvers based on algebraic domain decomposition techniques

One route to the parallel scalable solution of large sparse linear systems in parallel scientific computing is the use of hybrid methods that hierarchically combine direct and iterative methods. These techniques inherit the advantages of each approach, namely the limited amount of memory and natural parallelization for the iterative component and the numerical robustness of the direct part. The general underlying ideas are not new since they have been intensively used to design domain decomposition techniques; those approaches cover a fairly large range of computing techniques for the numerical solution of partial differential equations (PDEs) in time and space. Generally speaking, it refers to the splitting of the computational domain into sub-domains with or without overlap. The splitting strategy is generally governed by various constraints/objectives but the main one is to express parallelism. The numerical properties of the PDEs to be solved are usually intensively exploited at

the continuous or discrete levels to design the numerical algorithms so that the resulting specialized technique will only work for the class of linear systems associated with the targeted PDE.

In that context, we intend to continue our effort on the design of algebraic non-overlapping domain decomposition techniques that rely on the solution of a Schur complement system defined on the interface introduced by the partitioning of the adjacency graph of the sparse matrix associated with the linear system. Although it is better conditioned than the original system the Schur complement needs to be precondition to be amenable to a solution using a Krylov subspace method. Different hierarchical preconditioners will be considered, possibly multilevel, to improve the numerical behaviour of the current approaches implemented in our software libraries HIPS and MaPHyS. In addition to this numerical studies, advanced parallel implementation will be developed that will involve close collaborations between the hybrid and sparse direct activities. In particular some additional work to complete the initial study conducted with CEA-CESTA on full multigrid method will be undertaken. This activity will be developed either in the framework of the CEA-Inria agreement and/or through joint work with bresilian colleagues within the HOSCAR initiative.

3.3.3. Linear Krylov solvers

preconditioning is the main focus of the two activities described above. They aim at speeding up the convergence of a Krylov subspace method that is the complementary component involved in the solvers of interest for us. In that framework, we believe that various aspects deserve to be investigated; we will consider the following ones:

- preconditioned block Krylov solvers for multiple right-hand sides. In many large scientific and industrial applications, one has to solve a sequence of linear systems with several right-hand sides given simultaneously or in sequence (radar cross section calculation in electromagnetism, various source locations in seismic, parametric studies in general, ...). For “simultaneous" right-hand sides, the solvers of choice have been for years based on matrix factorizations as the factorization is performed once and simple and cheap block forward/backward substitutions are then performed. In order to effectively propose alternative to such solvers, we need to have efficient preconditioned Krylov subspace solvers. In that framework, block Krylov approaches, where the Krylov spaces associated with each right-hand side are shared to enlarge the search space will be considered. They are not only attractive because of this numerical feature (larger search space), but also from an implementation point of view. Their block-structures exhibit nice features with respect to data locality and re-usability that comply with the memory constraint of multicore architectures. Following the initial work by J. Yan Fei during his post-doc in HIEPACS, we will continue the numerical study of the block GMRES variant that combine inexact break-down detection and deflation at restart. In addition a special attention will be paid to situations where a massive number of right-hand sides are given where variants exploiting the possible sparsness (i.e., compression using H-matrix arithmetic) of these right-hand sides will be explored to design efficient numerical algorithms. Beyond new numerical investigations, a software implementation to be included in our linear solver libray will be developed.

  For right-hand sides available one after each other, various strategies that exploit the information available in the sequence of Krylov spaces (e.g. spectral information) will be considered that include for instance technique to perform incremental update of the preconditioner or to built augmented Krylov subspaces.
- Extension or modification of Krylov subspace algorithms for multicore architectures: finally to match as much as possible to the computer architecture evolution and get as much as possible performance out of the computer, a particular attention will be paid to adapt, extend or develop numerical schemes that comply with the efficiency constraints associated with the available computers. Nowadays, multicore architectures seem to become widely used, where memory latency and bandwidth are the main bottlenecks; investigations on communication avoiding techniques will be undertaken in the framework of preconditioned Krylov subspace solvers as a general guideline for all the items mentioned above. This research activity will benefit from the starting FP7 EXA2CT project led by HIEPACS on behalf of the IPL C2S@EXA that involves two other Inria projects namely ALPINES

and SAGE.

3.3.4. Eigensolvers

Many eigensolvers also rely on Krylov subspace techniques. Naturally some links exist between the Krylov subspace linear solvers and the Krylov subspace eigensolvers. We plan to study the computation of eigenvalue problems with respect to the following two different axes:

- Exploiting the link between Krylov subspace methods for linear system solution and eigensolvers, we intend to develop advanced iterative linear methods based on Krylov subspace methods that use some spectral information to build part of a subspace to be recycled, either though space augmentation or through preconditioner update. This spectral information may correspond to a certain part of the spectrum of the original large matrix or to some approximations of the eigenvalues obtained by solving a reduced eigenproblem. This technique will also be investigated in the framework of block Krylov subspace methods.
- In the context of the calculation of the ground state of an atomistic system, eigenvalue computation is a critical step; more accurate and more efficient parallel and scalable eigensolvers are required.

3.4. High performance Fast Multipole Method for N-body problems

Participants: Emmanuel Agullo, Bérenger Bramas, Arnaud Etcheverry, Olivier Coulaud, Matthias Messner, Cyrille Piacibello, Guillaume Sylvand.

In most scientific computing applications considered nowadays as computational challenges (like biological and material systems, astrophysics or electromagnetism), the introduction of hierarchical methods based on an octree structure has dramatically reduced the amount of computation needed to simulate those systems for a given accuracy. For instance, in the N-body problem arising from these application fields, we must compute all pairwise interactions among N objects (particles, lines, ...) at every timestep. Among these methods, the Fast Multipole Method (FMM) developed for gravitational potentials in astrophysics and for electrostatic (coulombic) potentials in molecular simulations solves this N-body problem for any given precision with $O(N)$ runtime complexity against $O(N^2)$ for the direct computation.

The potential field is decomposed in a near field part, directly computed, and a far field part approximated thanks to multipole and local expansions. We introduced a matrix formulation of the FMM that exploits the cache hierarchy on a processor through the Basic Linear Algebra Subprograms (BLAS). Moreover, we developed a parallel adaptive version of the FMM algorithm for heterogeneous particle distributions, which is very efficient on parallel clusters of SMP nodes. Finally on such computers, we developed the first hybrid MPI-thread algorithm, which enables to reach better parallel efficiency and better memory scalability. We plan to work on the following points in HIEPACS.

3.4.1. Improvement of calculation efficiency

Nowadays, the high performance computing community is examining alternative architectures that address the limitations of modern cache-based designs. GPU (Graphics Processing Units) and the Cell processor have thus already been used in astrophysics and in molecular dynamics. The Fast Mutipole Method has also been implemented on GPU. We intend to examine the potential of using these forthcoming processors as a building block for high-end parallel computing in N-body calculations. More precisely, we want to take advantage of our specific underlying BLAS routines to obtain an efficient and easily portable FMM for these new architectures. Algorithmic issues such as dynamic load balancing among heterogeneous cores will also have to be solved in order to gather all the available computation power. This research action will be conduced on close connection with the activity described in Section 3.2 .

3.4.2. Non uniform distributions

In many applications arising from material physics or astrophysics, the distribution of the data is highly non uniform and the data can grow between two time steps. As mentioned previously, we have proposed a hybrid MPI-thread algorithm to exploit the data locality within each node. We plan to further improve the load balancing for highly non uniform particle distributions with small computation grain thanks to dynamic load balancing at the thread level and thanks to a load balancing correction over several simulation time steps at the process level.

3.4.3. Fast multipole method for dislocation operators

The engine that we develop will be extended to new potentials arising from material physics such as those used in dislocation simulations. The interaction between dislocations is long ranged ($O(1/r)$) and anisotropic, leading to severe computational challenges for large-scale simulations. Several approaches based on the FMM or based on spatial decomposition in boxes are proposed to speed-up the computation. In dislocation codes, the calculation of the interaction forces between dislocations is still the most CPU time consuming. This computation has to be improved to obtain faster and more accurate simulations. Moreover, in such simulations, the number of dislocations grows while the phenomenon occurs and these dislocations are not uniformly distributed in the domain. This means that strategies to dynamically balance the computational load are crucial to achieve high performance.

3.4.4. Fast multipole method for boundary element methods

The boundary element method (BEM) is a well known solution of boundary value problems appearing in various fields of physics. With this approach, we only have to solve an integral equation on the boundary. This implies an interaction that decreases in space, but results in the solution of a dense linear system with $O(N^3)$ complexity. The FMM calculation that performs the matrix-vector product enables the use of Krylov subspace methods. Based on the parallel data distribution of the underlying octree implemented to perform the FMM, parallel preconditioners can be designed that exploit the local interaction matrices computed at the finest level of the octree. This research action will be conduced on close connection with the activity described in Section 3.3 . Following our earlier experience, we plan to first consider approximate inverse preconditionners that can efficiently exploit these data structures.

3.5. Efficient algorithmic for load balancing and code coupling in complex simulations

Participants: Astrid Casadei, Olivier Coulaud, Aurélien Esnard, Maria Predari, Pierre Ramet, Jean Roman, Clément Vuchener.

Many important physical phenomena in material physics and climatology are inherently complex applications. They often use multi-physics or multi-scale approaches, that couple different models and codes. The key idea is to reuse available legacy codes through a coupling framework instead of merging them into a standalone application. There is typically one model per different scale or physics; and each model is implemented by a parallel code. For instance, to model a crack propagation, one uses a molecular dynamic code to represent the atomistic scale and an elasticity code using a finite element method to represent the continuum scale. Indeed, fully microscopic simulations of most domains of interest are not computationally feasible. Combining such different scales or physics are still a challenge to reach high performance and scalability. If the model aspects are often well studied, there are several open algorithmic problems, that we plan to investigate in the HIEPACS project-team.

3.5.1. Efficient schemes for multiscale simulations

As mentioned previously, many important physical phenomena, such as material deformation and failure (see Section 4.1), are inherently multiscale processes that cannot always be modeled via continuum model. Fully microscopic simulations of most domains of interest are not computationally feasible. Therefore, researchers must look at multiscale methods that couple micro models and macro models. Combining different scales

such as quantum-atomistic or atomistic, mesoscale and continuum, are still a challenge to obtain efficient and accurate schemes that efficiently and effectively exchange information between the different scales. We are currently involved in two national research projects, that focus on multiscale schemes. More precisely, the models that we start to study are the quantum to atomic coupling (QM/MM coupling) in the ANR NOSSI and the atomic to dislocation coupling in the ANR OPTIDIS.

3.5.2. Dynamic load balancing for massively parallel coupled codes

In this context of code coupling, one crucial issue is undoubtedly the load balancing of the whole coupled simulation that remains an open question. The goal here is to find the best data distribution for the whole coupled simulation and not only for each standalone code, as it is most usually done. Indeed, the naive balancing of each code on its own can lead to an important imbalance and to a communication bottleneck during the coupling phase, that can drastically decrease the overall performance. Therefore, one argues that it is required to model the coupling itself in order to ensure a good scalability, especially when running on massively parallel architectures (tens of thousands of processors/cores). In other words, one must develop new algorithms and software implementation to perform a *coupling-aware* partitioning of the whole application.

Another related problem is the problem of resource allocation. This is particularly important for the global coupling efficiency and scalabilty, because each code involved in the coupling can be more or less computationally intensive, and there is a good trade-off to find between resources assigned to each code to avoid that one of them wait for the other(s). And what happens if the load of one code dynamically changes relatively to the other? In such a case, it could be convenient to dynamically adapt the number of resources used at runtime.

For instance, the conjugate heat transfer simulation in complex geometries (as developed by the CFD team of CERFACS) requires to couple a fluid/convection solver (AVBP) with a solid/conduction solver (AVTP). The AVBP code is much more CPU consuming than the AVTP code. As a consequence, there is an important computational imbalance between the two solvers. The use of new algorithms to correctly load balance coupled simulations with enhanced graph partitioning techniques appears as a promising way to reach better performances of coupled application on massively parallel computers.

3.5.3. Graph partitioning for hybrid solvers

Graph handling and partitioning play a central role in the activity described here but also in other numerical techniques detailed in Section 3.3 .

The Nested Dissection is now a well-known heuristic for sparse matrix ordering to both reduce the fill-in during numerical factorization and to maximize the number of independent computation tasks. By using the block data structure induced by the partition of separators of the original graph, very efficient parallel block solvers have been designed and implemented according to supernodal or multifrontal approaches. Considering hybrid methods mixing both direct and iterative solvers such as `HIPS` or `MaPHyS`, obtaining a domain decomposition leading to a good balancing of both the size of domain interiors and the size of interfaces is a key point for load balancing and efficiency in a parallel context. We intend to revisit some well-known graph partitioning techniques in the light of the hybrid solvers and design new algorithms to be tested in the `Scotch` package.

PHOENIX Project-Team

3. Research Program

3.1. Design-Driven Software Development

Raising the level of abstraction beyond programming is a very active research topic involving a range of areas, including software engineering, programming languages and formal verification. The challenge is to allow design dimensions of a software system, both functional and non-functional, to be expressed in a high-level way, instead of being encoded with a programming language. Such design dimensions can then be leveraged to verify conformance properties and to generate programming support.

Our research on this topic is to take up this challenge with an approach inspired by programming languages, introducing a full-fledged language for designing software systems and processing design descriptions both for verification and code generation purposes. Our approach is also DSL-inspired in that it defines a conceptual framework to guide software development. Lastly, to make our approach practical to software developers, we introduce a methodology and a suite of tools covering the development life-cycle.

To raise the level of abstraction beyond programming, the key approaches are model-driven engineering and architecture description languages. A number of *architecture description languages* have been proposed; they are either (1) coupled with a programming language (*e.g.,* [32]), providing some level of abstraction above programming, or (2) integrated into a programming language (*e.g.,* [26], [33]), mixing levels of abstraction. Furthermore, these approaches poorly leverage architecture descriptions to support programming, they are crudely integrated into existing development environments, or they are solely used for verification purposes. *Model-driven software development* is another actively researched area. This approach often lacks code generation and verification support. Finally, most (if not all) approaches related to our research goal are *general purpose*; their universal nature provides little, if any, guidance to design a software system. This situation is a major impediment to both reasoning about a design artifact and generating programming support.

3.2. Integrating Non-Functional Concerns into Software Design

Most existing design approaches do not address non-functional concerns. When they do, they do not provide an approach to non-functional concerns that covers the entire development life-cycle. Furthermore, they usually are general purpose, impeding the use of non-functional declarations for verification and code generation. For example, the Architecture Analysis & Design Language (AADL) is a standard dedicated to real-time embedded systems [28]. AADL provides language constructs for the specification of software systems (*e.g.,* component, port) and their deployment on execution platforms (*e.g.,* thread, process, memory). Using AADL, designers specify non-functional aspects by adding properties on language constructs (*e.g.,* the period of a thread) or using language extensions such as the Error Model Annex. [1] The software design concepts of AADL are still rather general purpose and give little guidance to the designer.

Beyond offering a conceptual framework, our language-based approach provides an ideal setting to address non-functional properties (*e.g.,* performance, reliability, security, ...). Specifically, a design language can be enriched with non-functional declarations to pursue two goals: (1) expanding further the type of conformance that can be checked between the design of a software system and its implementation, and (2) enabling additional programming support and guidance.

We are investigating this idea by extending our design language with non-functional declarations. For example, we have addressed error handling [10], access conflicts to resources [30], and quality of service constraints [29].

[1] The Error Model Annex is a standardized AADL extension for the description of errors [34].

Following our approach to paradigm-oriented software development, non-functional declarations are verified at design time, they generate support that guides and constrains programming, they produce a runtime system that preserves invariants.

3.3. Human-driven Software Design

Knowledge of the human characteristics (individual, social and organizational) allow the design of complex system and artifacts for increasing their efficacy. In our approach of assistive computing, a main challenge is the integration of facets of Human Factors in order to design technology support adapted to user needs in term of ergonomic properties (acceptability, usability, utility etc) and delivered functionalities (oriented task under user abilities contraints).

We adapt this approach to improve the independent living and self-determination of users with cognitive impairments by developing a variety of orchestration scenarios of networked objects (hardware/software) to provide a pervasive support to their activities. Human factors methodologies are adopted in our approach with the direct purpose the reliability and efficiency of the performance of digital support systems in respect of objectives of health and well-being of the person (monitoring, evaluation, and rehabilitation).

Precisely, our methodologies are based on a closed iterative loop, as described in the figure below :

- Identifying the person needs in a natural situation (i.e. , desired but problematic activities) according to Human Factors Models of activity (i.e., environmental constraints; social support networks - caregivers and family; person's abilities)
- Design pro- against - environmental measures that will being support the digital assistance to bypass cognitive impairment (ie , according to environmental models of cognitive compensatory mechanisms); and then develop solutions in terms of technological support (scenarios of networked objects, hardware interface, software interface , interaction style, etc)
- Empirical evaluation based on human experimentations that includes ergonomic assessments (acceptability , usability , usefulness, etc) as well as longitudinal evaluations of use's efficacy in terms of activities performed by the individual, of satisfaction and well -being provided to the individual but also to his/her entourage (family and caregivers).

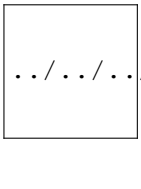


Figure 1. User-Centred Approach

RUNTIME Project-Team

3. Research Program

3.1. Runtime Systems Evolution

parallel,distributed,cluster,environment,library,communication,multithreading,multicore

This research project takes place within the context of high-performance computing. It seeks to contribute to the design and implementation of parallel runtime systems that shall serve as a basis for the implementation of high-level parallel middleware. Today, the implementation of such software (programming environments, numerical libraries, parallel language compilers, parallel virtual machines, etc.) has become so complex that the use of portable, low-level runtime systems is unavoidable.

Our research project centers on three main directions:

Mastering large, hierarchical multiprocessor machines With the beginning of the new century, computer makers have initiated a long term move of integrating more and more processing units, as an answer to the frequency wall hit by the technology. This integration cannot be made in a basic, planar scheme beyond a couple of processing units for scalability reasons. Instead, vendors have to resort to organize those processing units following some hierarchical structure scheme. A level in the hierarchy is then materialized by small groups of units sharing some common local cache or memory bank. Memory accesses outside the locality of the group are still possible thanks to bus-level consistency mechanisms but are significantly more expensive than local accesses, which, by definition, characterizes NUMA architectures.

Thus, the task scheduler must feed an increasing number of processing units with work to execute and data to process while keeping the rate of penalized memory accesses as low as possible. False sharing, ping-pong effects, data vs task locality mismatches, and even task vs task locality mismatches between tightly synchronizing activities are examples of the numerous sources of overhead that may arise if threads and data are not distributed properly by the scheduler. To avoid these pitfalls, the scheduler therefore needs accurate information both about the computing platform layout it is running on and about the structure and activities relationships of the application it is scheduling.

As quoted by Gao *et al.* [43], we believe it is important to expose domain-specific knowledge semantics to the various software components in order to organize computation according to the application and architecture. Indeed, the whole software stack, from the application to the scheduler, should be involved in the parallelizing, scheduling and locality adaptation decisions by providing useful information to the other components. Unfortunately, most operating systems only provide a poor scheduling API that does not allow applications to transmit valuable *hints* to the system.

This is why we investigate new approaches in the design of thread schedulers, focusing on high-level abstractions to both model hierarchical architectures and describe the structure of applications' parallelism. In particular, we have introduced the *bubble* scheduling concept [7] that helps to structure relations between threads in a way that can be efficiently exploited by the underlying thread scheduler. *Bubbles* express the inherent parallel structure of multithreaded applications: they are abstractions for grouping threads which "work together" in a recursive way. We are exploring how to dynamically schedule these irregular nested sets of threads on hierarchical machines [3], the key challenge being to schedule related threads as closely as possible in order to benefit from cache effects and avoid NUMA penalties. We are also exploring how to improve the transfer of scheduling hints from the programming environment to the runtime system, to achieve better computation efficiency.

This is also the reason why we explore new languages and compiler optimizations to better use domain specific information. We propose a new domain specific language, QIRAL, to generate parallel codes from high level formulations for Lattice QCD problems. QIRAL describes the formulation of the algorithms, of the matrices and preconditions used in this domain and generalizes languages such as SPIRAL used in auto-tuning library generator for signal processing applications. Lattice QCD applications require huge amount of processing power, on multinode, multi-core with GPUs. Simulation codes require to find new algorithms and efficient parallelization. So far, the difficulties for orchestrating parallelism efficiently hinder algorithmic exploration. The objective of QIRAL is to decouple algorithm exploration with parallelism description. Compiling QIRAL uses rewriting techniques for algorithm exploration, parallelization techniques for parallel code generation and potentially, runtime support to orchestrate this parallelism. Results of this work have been published in [9]. A similar approach, this time targeting methods to solve matrix equations, has been proposed [17]. Hydra focuses on systems of equations involving regular shaped matrices (such as upper triangular for instance) and finds automatically a parallel method to solve this system. The approach, using to a divide and conquer technique, works for several equations such as LU decomposition, Sylvester equation and has been shown to be comparable or outperforming Intel MKL library on multicores. Hydra relies on StarPU.

For parallel programs running on multicores, thread affinity and data locality is essential for performance. We investigated in [23] how thread pinning strategies could impact performance and performance stability and compared the efficiency of several profile-guided strategies with compile-time strategies. Following this effort, in MAQAO, we developped a language to ease the instrumentation of parallel codes, in particular for capturing memory traces [16]. Through the combined analysis of the code behavior, at compile time and at runtime, MAQAO can then help users to better pinpoint and quantify performance issues in OpenMP codes, find load imbalance between threads, size of working sets, false sharing situations... The MAQAO instrumentation language has been used successfully in other tools, such as TAU. Besides, we proposed in [15] to combine static and dynamic dependence analysis for the detection of vectorization opportunities. MAQAO then estimates the potential gain that could be reached through vectorization and identifies the required code transformations, either by changing loop control or data layout.

Aside from greedily invading all these new cores, demanding HPC applications now throw excited glances at the appealing computing power left unharvested inside the graphical processing units (GPUs). A strong demand is arising from the application programmers to be given means to access this power without bearing an unaffordable burden on the portability side. Efforts have already been made by the community in this respect but the tools provided still are rather close to the hardware, if not to the metal. Hence, we decided to launch some investigations on addressing this issue. In particular, we have designed a programming environment named StarPU that enables the programmer to offload tasks onto such heterogeneous processing units and gives that programmer tools to fit tasks to processing units capability, tools to efficiently manage data moves to and from the offloading hardware and handles the scheduling of such tasks all in an abstracted, portable manner. The challenge here is to take into account the intricacies of all computation unit: not only the computation power is heterogeneous among the machine, but data transfers themselves have various behavior depending on the machine architecture and GPUs capabilities, and thus have to be taken into account to get the best performance from the underlying machine. As a consequence, StarPU not only pays attention to fully exploit each of the different computational resources at the same time by properly mapping tasks in a dynamic manner according to their computation power and task behavior by the means of scheduling policies, but it also provides a distributed shared-memory library that makes it possible to manipulate data across heterogeneous multicore architectures in a high-level fashion while being optimized according to the machine possibilities. In addition to this, the scheduling policy of StarPU has been modularized; this makes it easy to experiment with state of the art theoretical scheduling strategies. Last but not least, StarPU works over clusters, by extending the shared-memory view over the MPI communication library. This allows, with the same

sequential-looking application source code, to tackle all architectures from small multicore systems to clusters of heterogeneous systems.

We extended OpenCL capabilities by proposing to use, transparently, StarPU as an OpenCL device [35]. A functional approach to StarPU has been proposed besides in [18].

Optimizing communications over high performance clusters and grids Using a large panel of mechanisms such as user-mode communications, zero-copy transactions and communication operation offload, the critical path in sending and receiving a packet over high speed networks has been drastically reduced over the years. Recent implementations of the MPI standard, which have been carefully designed to directly map *basic* point-to-point requests onto the underlying low-level interfaces, almost reach the same level of performance for very basic point-to-point messaging requests. However more complex requests such as non-contiguous messages are left mostly unattended, and even more so are the irregular and multiflow communication schemes. The intent of the work on our NewMadeleine communication engine, for instance, is to address this situation thoroughly. The NewMadeleine optimization layer delivers much better performance on *complex* communication schemes with negligible overhead on basic single packet point-to-point requests. Through Mad-MPI, our proof-of-concept implementation of a subset of the MPI API, we intend to show that MPI applications can also benefit from the NewMadeleine communication engine.

The increasing number of cores in cluster nodes also raises the importance of intra-node communication. Our KNem software module aims at offering optimized communication strategies for this special case and let the above MPI implementations benefit from dedicated models depending on process placement and hardware characteristics.

Moreover, the convergence between specialized high-speed networks and traditional Ethernet networks leads to the need to adapt former software and hardware innovations to new message-passing stacks. Our work on the Open-MX software is carried out in this context.

Regarding larger scale configurations (clusters of clusters, grids), we intend to propose new models, principles and mechanisms that should allow to combine communication handling, threads scheduling and I/O event monitoring on such architectures, both in a portable and efficient way. We particularly intend to study the introduction of new runtime system functionalities to ease the development of code-coupling distributed applications, while minimizing their unavoidable negative impact on the application performance.

Integrating Communications and Multithreading Asynchronism is becoming ubiquitous in modern communication runtimes. Complex optimizations based on online analysis of the communication schemes and on the de-coupling of the request submission vs processing. Flow multiplexing or transparent heterogeneous networking also imply an active role of the runtime system request submit and process. And communication overlap as well as reactiveness are critical. Since network request cost is in the order of magnitude of several thousands CPU cycles at least, independent computations should not get blocked by an ongoing network transaction. This is even more true with the increasingly dense SMP, multicore, SMT architectures where many computing units share a few NICs. Since portability is one of the most important requirements for communication runtime systems, the usual approach to implement asynchronous processing is to use threads (such as Posix threads). Popular communication runtimes indeed are starting to make use of threads internally and also allow applications to also be multithreaded. Low level communication libraries also make use of multithreading. Such an introduction of threads inside communication subsystems is not going without troubles however. The fact that multithreading is still usually optional with these runtimes is symptomatic of the difficulty to get the benefits of multithreading in the context of networking without suffering from the potential drawbacks. We advocate the importance of the cooperation between the asynchronous event management code and the thread scheduling code in order to avoid such disadvantages. We intend to propose a framework for symbiotically combining both approaches inside a new generic I/O event manager.

Moreover, the design of distributed parallel code, integrating both MPI and OpenMP, is complex and error-prine. Deadlock situations may arise and are difficult to detect. We proposed an original approach, based on static (compile-time) analysis and runtime verification in order to detect deadlock situation but also to pinpoint the cause of such deadlock [27]. This work first focuses on MPI communication alone, the extension to hybrid MPI/OpenMP codes is in progress.

FLOWERS Project-Team

3. Research Program

3.1. Research Program

Research in artificial intelligence, machine learning and pattern recognition has produced a tremendous amount of results and concepts in the last decades. A blooming number of learning paradigms - supervised, unsupervised, reinforcement, active, associative, symbolic, connectionist, situated, hybrid, distributed learning... - nourished the elaboration of highly sophisticated algorithms for tasks such as visual object recognition, speech recognition, robot walking, grasping or navigation, the prediction of stock prices, the evaluation of risk for insurances, adaptive data routing on the internet, etc... Yet, we are still very far from being able to build machines capable of adapting to the physical and social environment with the flexibility, robustness, and versatility of a one-year-old human child.

Indeed, one striking characteristic of human children is the nearly open-ended diversity of the skills they learn. They not only can improve existing skills, but also continuously learn new ones. If evolution certainly provided them with specific pre-wiring for certain activities such as feeding or visual object tracking, evidence shows that there are also numerous skills that they learn smoothly but could not be "anticipated" by biological evolution, for example learning to drive a tricycle, using an electronic piano toy or using a video game joystick. On the contrary, existing learning machines, and robots in particular, are typically only able to learn a single pre-specified task or a single kind of skill. Once this task is learnt, for example walking with two legs, learning is over. If one wants the robot to learn a second task, for example grasping objects in its visual field, then an engineer needs to re-program manually its learning structures: traditional approaches to task-specific machine/robot learning typically include engineer choices of the relevant sensorimotor channels, specific design of the reward function, choices about when learning begins and ends, and what learning algorithms and associated parameters shall be optimized.

As can be seen, this requires a lot of important choices from the engineer, and one could hardly use the term "autonomous" learning. On the contrary, human children do not learn following anything looking like that process, at least during their very first years. Babies develop and explore the world by themselves, focusing their interest on various activities driven both by internal motives and social guidance from adults who only have a folk understanding of their brains. Adults provide learning opportunities and scaffolding, but eventually young babies always decide for themselves what activity to practice or not. Specific tasks are rarely imposed to them. Yet, they steadily discover and learn how to use their body as well as its relationships with the physical and social environment. Also, the spectrum of skills that they learn continuously expands in an organized manner: they undergo a developmental trajectory in which simple skills are learnt first, and skills of progressively increasing complexity are subsequently learnt.

A link can be made to educational systems where research in several domains have tried to study how to provide a good learning experience to learners. This includes the experiences that allow better learning, and in which sequence they must be experienced. This problem is complementary to that of the learner that tries to learn efficiently, and the teacher here has to use as efficienlty the limited time and motivational resources of the learner. Several results from psychology [70] and neuroscience [10] have argued that the human brain feels intrinsic pleasure in practicing activities of optimal difficulty or challenge. A teacher must exploit such activities to create positive psychological states of flow [76].

A grand challenge is thus to be able to build robotic machines that possess this capability to discover, adapt and develop continuously new know-how and new knowledge in unknown and changing environments, like human children. In 1950, Turing wrote that the child's brain would show us the way to intelligence: "Instead of trying to produce a program to simulate the adult mind, why not rather try to produce one which simulates the child's" [133]. Maybe, in opposition to work in the field of Artificial Intelligence who has focused on mechanisms trying to match the capabilities of "intelligent" human adults such as chess playing or natural language

dialogue [91], it is time to take the advice of Turing seriously. This is what a new field, called developmental (or epigenetic) robotics, is trying to achieve [100] [135]. The approach of developmental robotics consists in importing and implementing concepts and mechanisms from developmental psychology [107], cognitive linguistics [75], and developmental cognitive neuroscience [95] where there has been a considerable amount of research and theories to understand and explain how children learn and develop. A number of general principles are underlying this research agenda: embodiment [72] [119], grounding [89], situatedness [66], self-organization [131] [120], enaction [134], and incremental learning [73].

Among the many issues and challenges of developmental robotics, two of them are of paramount importance: exploration mechanisms and mechanisms for abstracting and making sense of initially unknown sensorimotor channels. Indeed, the typical space of sensorimotor skills that can be encountered and learnt by a developmental robot, as those encountered by human infants, is immensely vast and inhomogeneous. With a sufficiently rich environment and multimodal set of sensors and effectors, the space of possible sensorimotor activities is simply too large to be explored exhaustively in any robot's life time: it is impossible to learn all possible skills and represent all conceivable sensory percepts. Moreover, some skills are very basic to learn, some other very complicated, and many of them require the mastery of others in order to be learnt. For example, learning to manipulate a piano toy requires first to know how to move one's hand to reach the piano and how to touch specific parts of the toy with the fingers. And knowing how to move the hand might require to know how to track it visually.

Exploring such a space of skills randomly is bound to fail or result at best on very inefficient learning [15]. Thus, exploration needs to be organized and guided. The approach of epigenetic robotics is to take inspiration from the mechanisms that allow human infants to be progressively guided, i.e. to develop. There are two broad classes of guiding mechanisms which control exploration:

1. **internal guiding mechanisms,** and in particular intrinsic motivation, responsible of spontaneous exploration and curiosity in humans, which is one of the central mechanisms investigated in FLOWERS, and technically amounts to achieve online active self-regulation of the growth of complexity in learning situations;
2. **social learning and guidance,** a learning mechanisms that exploits the knowledge of other agents in the environment and/or that is guided by those same agents. These mechanisms exist in many different forms like emotional reinforcement, stimulus enhancement, social motivation, guidance, feedback or imitation, some of which being also investigated in FLOWERS;

3.1.1. Internal guiding mechanisms

In infant development, one observes a progressive increase of the complexity of activities with an associated progressive increase of capabilities [107], children do not learn everything at one time: for example, they first learn to roll over, then to crawl and sit, and only when these skills are operational, they begin to learn how to stand. The perceptual system also gradually develops, increasing children perceptual capabilities other time while they engage in activities like throwing or manipulating objects. This make it possible to learn to identify objects in more and more complex situations and to learn more and more of their physical characteristics.

Development is therefore progressive and incremental, and this might be a crucial feature explaining the efficiency with which children explore and learn so fast. Taking inspiration from these observations, some roboticists and researchers in machine learning have argued that learning a given task could be made much easier for a robot if it followed a developmental sequence and "started simple" [68] [80]. However, in these experiments, the developmental sequence was crafted by hand: roboticists manually build simpler versions of a complex task and put the robot successively in versions of the task of increasing complexity. And when they wanted the robot to learn a new task, they had to design a novel reward function.

Thus, there is a need for mechanisms that allow the autonomous control and generation of the developmental trajectory. Psychologists have proposed that intrinsic motivations play a crucial role. Intrinsic motivations are mechanisms that push humans to explore activities or situations that have intermediate/optimal levels of novelty, cognitive dissonance, or challenge [70] [76] [79]. The role and structure of intrinsic motivation in humans have been made more precise thanks to recent discoveries in neuroscience showing the implication

of dopaminergic circuits and in exploration behaviors and curiosity [78] [92] [125]. Based on this, a number of researchers have began in the past few years to build computational implementation of intrinsic motivation [15] [117] [123] [69] [93] [105] [124]. While initial models were developed for simple simulated worlds, a current challenge is to manage to build intrinsic motivation systems that can efficiently drive exploratory behaviour in high-dimensional unprepared real world robotic sensorimotor spaces [117][15] [118] [122]. Specific and complex problems are posed by real sensorimotor spaces, in particular due to the fact that they are both high-dimensional as well as (usually) deeply inhomogeneous. As an example for the latter issue, some regions of real sensorimotor spaces are often unlearnable due to inherent stochasticity or difficulty, in which case heuristics based on the incentive to explore zones of maximal unpredictability or uncertainty, which are often used in the field of active learning [74] [90] typically lead to catastrophic results. The issue of high dimensionality does not only concern motor spaces, but also sensory spaces, leading to the problem of correctly identifying, among typically thousands of quantities, those latent variables that have links to behavioral choices. In FLOWERS, we aim at developing intrinsically motivated exploration mechanisms that scale in those spaces, by studying suitable abstraction processes in conjunction with exploration strategies.

3.1.2. Socially Guided and Interactive Learning

Social guidance is as important as intrinsic motivation in the cognitive development of human babies [107]. There is a vast literature on learning by demonstration in robots where the actions of humans in the environment are recognized and transferred to robots [67]. Most such approaches are completely passive: the human executes actions and the robot learns from the acquired data. Recently, the notion of interactive learning has been introduced in [132], [71], motivated by the various mechanisms that allow humans to socially guide a robot [121]. In an interactive context the steps of self-exploration and social guidances are not separated and a robot learns by self exploration and by receiving extra feedback from the social context [132], [96] [106].

Social guidance is also particularly important for learning to segment and categorize the perceptual space. Indeed, parents interact a lot with infants, for example teaching them to recognize and name objects or characteristics of these objects. Their role is particularly important in directing the infant attention towards objects of interest that will make it possible to simplify at first the perceptual space by pointing out a segment of the environment that can be isolated, named and acted upon. These interactions will then be complemented by the children own experiments on the objects chosen according to intrinsic motivation in order to improve the knowledge of the object, its physical properties and the actions that could be performed with it.

In FLOWERS, we are aiming at including intrinsic motivation system in the self-exploration part thus combining efficient self-learning with social guidance [109], [110]. We also work on developing perceptual capabilities by gradually segmenting the perceptual space and identifying objects and their characteristics through interaction with the user [102] and robots experiments [94]. Another challenge is to allow for more flexible interaction protocols with the user in terms of what type of feedback is provided and how it is provided [98].

MANAO Team

3. Research Program

3.1. Related Scientific Domains

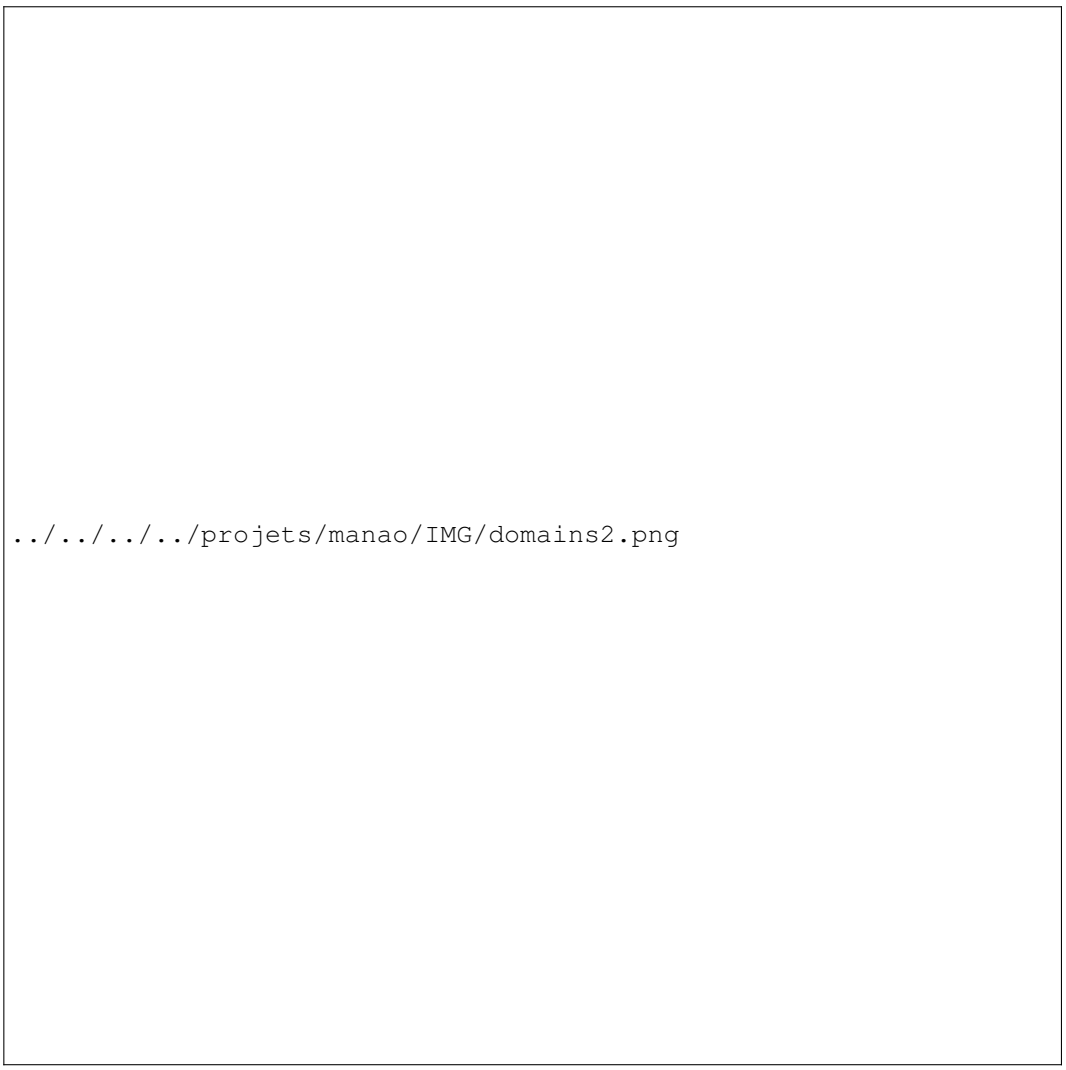


Figure 4. Related scientific domains of the MANAO project.

The *MANAO* project aims to study, acquire, model, and render the interactions between the three components that are light, shape, and matter from the viewpoint of an observer. As detailed more lengthily in the next section, such a work will be done using the following approach: first, we will tend to consider that these three components do not have strict frontiers when considering their impacts on the final observers; then, we will

not only work in **computer graphics**, but also at the intersections of computer graphics and **optics**, exploring the mutual benefits that the two domains may provide. It is thus intrinsically a **transdisciplinary** project (as illustrated in Figure 4) and we expect results in both domains.

Thus, the proposed team-project aims at establishing a close collaboration between computer graphics (e.g., 3D modeling, geometry processing, shading techniques, vector graphics, and GPU programming) and optics (e.g., design of optical instruments, and theories of light propagation). The following examples illustrate the strengths of such a partnership. First, in addition to simpler radiative transfer equations [49] commonly used in computer graphics, research in the later will be based on state-of-the-art understanding of light propagation and scattering in real environments. Furthermore, research will rely on appropriate instrumentation expertise for the measurement [61], [62] and display [60] of the different phenomena. Reciprocally, optics researches may benefit from the expertise of computer graphics scientists on efficient processing to investigate interactive simulation, visualization, and design. Furthermore, new systems may be developed by unifying optical and digital processing capabilities. Currently, the scientific background of most of the team members is related to computer graphics and computer vision. A large part of their work have been focused on simulating and analyzing optical phenomena as well as in acquiring and visualizing them. Combined with the close collaboration with the optics laboratory (LP2N) and with the students issued from the "Institut d'Optique", this background ensures that we can expect the following results from the project: the construction of a common vocabulary for tightening the collaboration between the two scientific domains and creating new research topics. By creating this context, we expect to attract (and even train) more trans-disciplinary researchers.

At the boundaries of the *MANAO* project lie issues in **human and machine vision**. We have to deal with the former whenever a human observer is taken into account. On one side, computational models of human vision are likely to guide the design of our algorithms. On the other side, the study of interactions between light, shape, and matter may shed some light on the understanding of visual perception. The same kind of connections are expected with machine vision. On the one hand, traditional computational methods for acquisition (such as photogrammetry) are going to be part of our toolbox. On the other hand, new display technologies (such as augmented reality) are likely to benefit from our integrated approach and systems. In the *MANAO* project we are mostly users of results from human vision. When required, some experimentation might be done in collaboration with experts from this domain, like with the European PRISM project (cf. Section 7.3). For machine vision, provided the tight collaboration between optical and digital systems, research will be carried out inside the *MANAO* project.

Analysis and modeling rely on **tools from applied mathematics** such as differential and projective geometry, multi-scale models, frequency analysis [51] or differential analysis [83], linear and non-linear approximation techniques, stochastic and deterministic integrations, and linear algebra. We not only rely on classical tools, but also investigate and adapt recent techniques (e.g., improvements in approximation techniques), focusing on their ability to run on modern hardware: the development of our own tools (such as Eigen, see Section 4.1) is essential to control their performances and their abilities to be integrated into real-time solutions or into new instruments.

3.2. Research axes

The *MANAO* project is organized around four research axes that cover the large range of expertise of its members and associated members. We briefly introduce these four axes in this section. More details and their inter-influences that are illustrated in the Figure 2 will be given in the following sections.

Axis 1 is the theoretical foundation of the project. Its main goal is to increase the understanding of light, shape, and matter interactions by combining expertise from different domains: optics and human/machine vision for the analysis and computer graphics for the simulation aspect. The goal of our analyses is to identify the different layers/phenomena that compose the observed signal. In a second step, the development of physical simulations and numerical models of these identified phenomena is a way to validate the pertinence of the proposed decompositions.

In Axis 2, the final observers are mainly physical captors. Our goal is thus the development of new acquisition and display technologies that combine optical and digital processes in order to reach fast transfers between real and digital worlds, in order to increase the convergence of these two worlds.

Axes 3 and 4 focus on two aspects of computer graphics: rendering, visualization and illustration in Axis 3, and editing and modeling (content creation) in Axis 4. In these two axes, the final observers are mainly human users, either generic users or expert ones (e.g., archaeologist [7], computer graphics artists).

3.3. Axis 1: Analysis and Simulation

Challenge: Definition and understanding of phenomena resulting from interactions between light, shape, and matter as seen from an observer point of view.

Results: Theoretical tools and numerical models for analyzing and simulating the observed optical phenomena.

To reach the goals of the *MANAO* project, we need to **increase our understanding** of how light, shape, and matter act together in synergy and how the resulting signal is finally observed. For this purpose, we need to identify the different phenomena that may be captured by the targeted observers. This is the main objective of this research axis, and it is achieved by using three approaches: the simulation of interactions between light, shape, and matter, their analysis and the development of new numerical models. This resulting improved knowledge is a foundation for the researches done in the three other axes, and the simulation tools together with the numerical models serve the development of the joint optical/digital systems in Axis 2 and their validation.

One of the main and earliest goals in computer graphics is to faithfully reproduce the real world, focusing mainly on light transport. Compared to researchers in physics, researchers in computer graphics rely on a subset of physical laws (mostly radiative transfer and geometric optics), and their main concern is to efficiently use the limited available computational resources while developing as fast as possible algorithms. For this purpose, a large set of tools has been introduced to take a **maximum benefit of hardware** specificities. These tools are often dedicated to specific phenomena (e.g., direct or indirect lighting, color bleeding, shadows, caustics). An efficiency-driven approach needs such a classification of light paths [57] in order to develop tailored strategies [100]. For instance, starting from simple direct lighting, more complex phenomena have been progressively introduced: first diffuse indirect illumination [55], [91], then more generic inter-reflections [64], [49] and volumetric scattering [88], [46]. Thanks to this search for efficiency and this classification, researchers in computer graphics have developed a now recognized expertise in fast-simulation of light propagation. Based on finite elements (radiosity techniques) or on unbiased Monte Carlo integration schemes (ray-tracing, particle-tracing, ...), the resulting algorithms and their combination are now sufficiently accurate to be used-back in physical simulations. The *MANAO* project will continue the search for **efficient and accurate simulation** techniques, but extending it from computer graphics to optics. Thanks to the close collaboration with scientific researchers from optics, new phenomena beyond radiative transfer and geometric optics will be explored.

Search for algorithmic efficiency and accuracy has to be done in parallel with **numerical models**. The goal of visual fidelity (generalized to accuracy from an observer point of view in the project) combined with the goal of efficiency leads to the development of alternative representations. For instance, common classical finite-element techniques compute only basis coefficients for each discretization element: the required discretization density would be too large and to computationally expensive to obtain detailed spatial variations and thus visual fidelity. Examples includes texture for decorrelating surface details from surface geometry and high-order wavelets for a multi-scale representation of lighting [45]. The numerical complexity explodes when considering directional properties of light transport such as radiance intensity (Watt per square meter and per steradian - $W.m^{-2}.sr^{-1}$), reducing the possibility to simulate or accurately represent some optical phenomena. For instance, Haar wavelets have been extended to the spherical domain [90] but are difficult to extend to non-piecewise-constant data [93]. More recently, researches prefer the use of Spherical Radial Basis Functions [97] or Spherical Harmonics [82]. For more complex data, such as reflective properties (e.g., BRDF [76], [65] - 4D), ray-space (e.g., Light-Field [73] - 4D), spatially varying reflective properties (6D

- [86]), new models, and representations are still investigated such as rational functions [79] or dedicated models [33] and parameterizations [89], [94]. For each (newly) defined phenomena, we thus explore the space of possible numerical representations to determine the **most suited one for a given application**, like we have done for BRDF [79].

Texuring | 1st order gradient field | Environment reflection | 2st order gradient field

Figure 5. First-oder analysis [102] have shown that shading variations are caused by depth variations (first-order gradient field) and by normal variations (second-order fields). These fields are visualized using hue and saturation to indicate direction and magnitude of the flow respectively.

Before being able to simulate or to represent the different **observed phenomena**, we need to define and describe them. To understand the difference between an observed phenomenon and the classical light, shape, and matter decomposition, we can take the example of a highlight. Its observed shape (by a human user or a sensor) is the resulting process of the interaction of these three components, and can be simulated this way. However, this does not provide any intuitive understanding of their relative influence on the final shape: an artist will directly describe the resulting shape, and not each of the three properties. We thus want to decompose the observed signal into models for each scale that can be easily understandable, representable, and manipulable. For this purpose, we will rely on the **analysis** of the resulting interaction of light, shape, and matter as observed by a human or a physical sensor. We first consider this analysis from an **optical point of view**, trying to identify the different phenomena and their scale according to their mathematical properties (e.g., differential [83] and frequency analysis [51]). Such an approach has leaded us to exhibit the influence of surfaces flows (depth and normal gradients) into lighting pattern deformation (see Figure 5). For a **human observer**, this correspond to one recent trend in computer graphics that takes into account the human visual systems [52] both to evaluate the results and to guide the simulations.

3.4. Axis 2: From Acquisition to Display

Challenge: Convergence of optical and digital systems to blend real and virtual worlds.

Results: Instruments to acquire real world, to display virtual world, and to make both of them interact.

For this axis, we investigate *unified acquisition and display systems*, that is systems which combine optical instruments with digital processing. From digital to real, we investigate new display approaches [73], [60]. We consider projecting systems and surfaces [41], for personal use, virtual reality and augmented reality [37]. From the real world to the digital world, we favor direct measurements of parameters for models and representations, using (new) optical systems unless digitization is required [54], [53]. These resulting systems have to acquire the different phenomena described in Axis 1 and to display them, in an efficient manner [58], [34], [59], [62]. By efficient, we mean that we want to shorten the path between the real world and the virtual world by increasing the data bandwidth between the real (analog) and the virtual (digital) worlds, and by reducing the latency for real-time interactions (we have to prevent unnecessary conversions, and to reduce

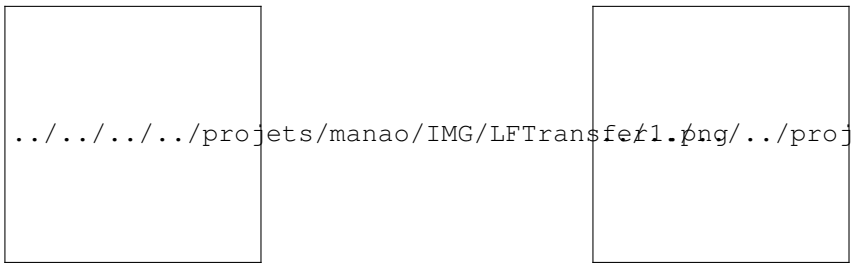


Figure 6. Light-Field transfer: global illumination between real and synthetic objects [44]

processing time). To reach this goal, the systems have to be designed as a whole, not by a simple concatenation of optical systems and digital processes, nor by considering each component independently [63].

To increase data bandwidth, one solution is to **parallelize more and more the physical systems**. One possible solution is to multiply the number of simultaneous acquisitions (e.g., simultaneous images from multiple viewpoints [62], [81]). Similarly, increasing the number of viewpoints is a way toward the creation of full 3D displays [73]. However, full acquisition or display of 3D real environments theoretically requires a continuous field of viewpoints, leading to huge data size. Despite the current belief that the increase of computational power will fill the missing gap, when it comes to visual or physical realism, if you double the processing power, people may want four times more accuracy, thus increasing data size as well. Furthermore, this leads to solutions that are not energy efficient and thus cannot be embedded into mobile devices. To reach the best performances, a trade-off has to be found between the amount of data required to represent accurately the reality and the amount of required processing. This trade-off may be achieved using **compressive sensing**. Compressive sensing is a new trend issued from the applied mathematics community that provides tools to accurately reconstruct a signal from a small set of measurements assuming that it is sparse in a transform domain (e.g., [80], [104]).

We prefer to achieve this goal by avoiding as much as possible the classical approach where acquisition is followed by a fitting step: this requires in general a large amount of measurements and the fitting itself may consume consequently too much memory and preprocessing time. By **preventing unnecessary conversion** through fitting techniques, such an approach increase the speed and reduce the data transfer for acquisition but also for display. One of the best recent examples is the work of Cossairt et al. [44]. The whole system is designed around a unique representation of the energy-field issued from (or leaving) a 3D object, either virtual or real: the Light-Field. A Light-Field encodes the light emitted in any direction from any position on an object. It is acquired thanks to a lens-array that leads to the capture of, and projection from, multiple simultaneous viewpoints. A unique representation is used for all the steps of this system. Lens-arrays, parallax barriers, and coded-aperture [69] are one of the key technologies to develop such acquisition (e.g., Light-Field camera [1] [63] and acquisition of light-sources [54]), projection systems (e.g., auto-stereoscopic displays). Such an approach is versatile and may be applied to improve classical optical instruments [68]. More generally, by designing unified optical and digital systems [77], it is possible to leverage the requirement of processing power, the memory footprint, and the cost of optical instruments.

Those are only some examples of what we investigate. We also consider the following approaches to develop new unified systems. First, similar to (and based on) the analysis goal of Axis 1, we have to take into account as much as possible the characteristics of the measurement setup. For instance, when fitting cannot be avoided, integrating them may improve both the processing efficiency and accuracy [79]. Second, we have to integrate signals from multiple sensors (such as GPS, accelerometer, ...) to prevent some computation (e.g., [70]).

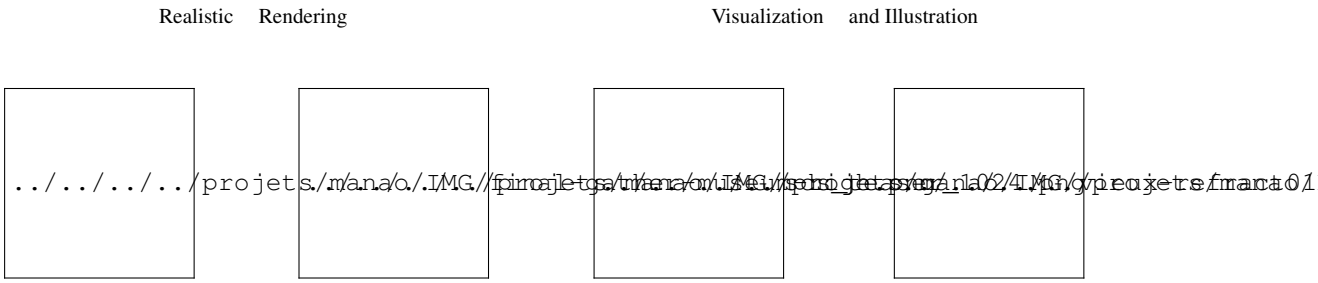
[1] Lytro, http://www.lytro.com/

Finally, the experience of the group in surface modeling help the design of optical surfaces [66] for light sources or head-mounted displays.

3.5. Axis 3: Rendering, Visualization and Illustration

Challenge: How to offer the most legible signal to the final observer in real-time?

Results: High-level shading primitives, expressive rendering techniques for object depiction, real-time realistic rendering algorithms



(a) Global illumination [78] (b) Shadows [35] (c) Shape enhancement [9] (d) Shape depiction [32]

Figure 7. In the MANAO project, we are investigating rendering techniques from realistic solutions (e.g., inter-reflections (a) and shadows (b)) to more expressive ones (shape enhancement (c) with realistic style and shape depiction (d) with stylized style) for visualization.

The main goal of this axis is to offer to the final observer, in this case mostly a human user, the most legible signal in real-time. Thanks to the analysis and to the decomposition in different phenomena resulting from interactions between light, shape, and matter (Axis 1), and their perception, we can use them to convey essential information in the most pertinent way. Here, the word *pertinent* can take various forms depending on the application.

In the context of scientific illustration and visualization, we are primarily interested in tools to convey shape or material characteristics of objects in animated 3D scenes. **Expressive rendering** techniques (see Figure 7 c,d) provide means for users to depict such features with their own style. To introduce our approach, we detail it from a shape-depiction point of view, domain where we have acquired a recognized expertise. Prior work in this area mostly focused on stylization primitives to achieve line-based rendering [10], [67] or stylized shading [39],[9] with various levels of abstraction. A clear representation of important 3D **object features** remains a major challenge for better shape depiction, stylization and abstraction purposes. Most existing representations provide only local properties (e.g., curvature), and thus lack characterization of broader shape features. To overcome this limitation, we are developing higher level descriptions of shape [31] with increased robustness to sparsity, noise, and outliers. This is achieved in close collaboration with Axis 1 by the use of higher-order local fitting methods, multi-scale analysis, and global regularization techniques. In order not to neglect the observer and the material characteristics of the objects, we couple this approach with an analysis of the appearance model. To our knowledge, this is an approach which has not been considered yet. This research direction is at the heart of the *MANAO* project, and has a strong connection with the analysis we plan to conduct in Axis 1. Material characteristics are always considered at the light ray level, but an understanding of **higher-level primitives** (like the shape of highlights and their motion) would help us to produce more legible renderings and permit novel stylizations; for instance, there is no method that is today able to create stylized renderings that follow the motion of highlights or shadows. We also believe such tools also play a fundamental role for geometry processing purposes (such as shape matching, reassembly, simplification), as well as for editing purposes as discussed in Axis 4.

In the context of **real-time photo-realistic rendering** ((see Figure 7 a,b), the challenge is to compute the most plausible images with minimal effort. During the last decade, a lot of work has been devoted to design approximate but real-time rendering algorithms of complex lighting phenomena such as soft-shadows [103], motion blur [51], depth of field [92], reflexions, refractions, and inter-reflexions. For most of these effects it becomes harder to discover fundamentally new and faster methods. On the other hand, we believe that significant speedup can still be achieved through more clever use of **massively parallel architectures** of the current and upcoming hardware, and/or through more clever tuning of the current algorithms. In particular, regarding the second aspect, we remark that most of the proposed algorithms depend on several parameters which can be used to **trade the speed over the quality**. Significant speed-up could thus be achieved by identifying effects that would be masked or facilitated and thus devote appropriate computational resources to the rendering [4], [50]. Indeed, the algorithm parameters controlling the quality vs speed are numerous without a direct mapping between their values and their effect. Moreover, their ideal values vary over space and time, and to be effective such an auto-tuning mechanism has to be extremely fast such that its cost is largely compensated by its gain. We believe that our various work on the analysis of the appearance such as in Axis 1 could be beneficial for such purpose too.

Realistic and real-time rendering is closely related to Axis 2: real-time rendering is a requirement to close the loop between real world and digital world. We have to thus develop algorithms and rendering primitives that allow the integration of the acquired data into real-time techniques. We have also to take care of that these real-time techniques have to work with new display systems. For instance, stereo, and more generally multi-view displays are based on the multiplication of simultaneous images. Brute force solutions consist in independent rendering pipeline for each viewpoint. A more energy-efficient solution would take advantages of the computation parts that may be factorized. Another example is the rendering techniques based on image processing, such as our work on augmented reality [43]. Independent image processing for each viewpoint may disturb the feeling of depth by introducing inconsistent information in each images. Finally, more dedicated displays [60] would require new rendering pipelines.

3.6. Axis 4: Editing and Modeling

Challenge: Editing and modeling appearance using drawing- or sculpting-like tools through high level representations.

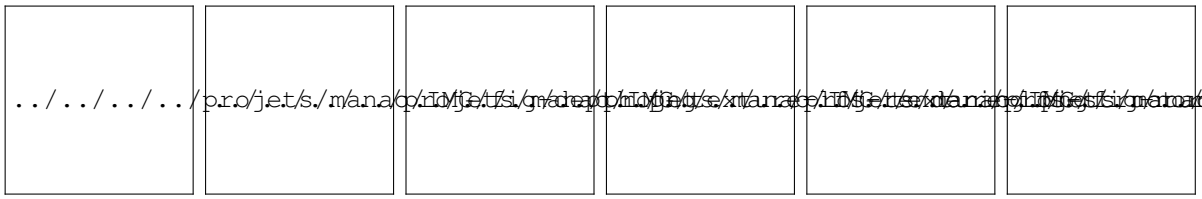
Results: High-level primitives and hybrid representations for appearance and shape.

During the last decade, the domain of computer graphics has exhibited tremendous improvements in image quality, both for 2D applications and 3D engines. This is mainly due to the availability of an ever increasing amount of shape details, and sophisticated appearance effects including complex lighting environments. Unfortunately, with such a growth in visual richness, even so-called *vectorial* representations (e.g., subdivision surfaces, Bézier curves, gradient meshes, etc.) become very dense and unmanageable for the end user who has to deal with a huge mass of control points, color labels, and other parameters. This is becoming a major challenge, with a necessity for novel representations. This Axis is thus complementary of Axis 3: the focus is the development of primitives that are easy to use for modeling and editing.

More specifically, we plan to investigate *vectorial representations* that would be amenable to the production of rich shapes with a minimal set of primitives and/or parameters. To this end we plan to build upon our insights on dynamic local reconstruction techniques and implicit surfaces [6], [1]. When working in 3D, an interesting approach to produce detailed shapes is by means of procedural geometry generation. For instance, many natural phenomena like waves or clouds may be modeled using a combination of procedural functions. Turning such functions into triangle meshes (main rendering primitives of GPUs) is a tedious process that appears not to be necessary with an adapted vectorial shape representation where one could directly turn procedural functions into implicit geometric primitives. Since we want to prevent unnecessary conversions in the whole pipeline (here, between modeling and rendering steps), we will also consider *hybrid representations* mixing meshes and implicit representations. Such research has thus to be conducted while considering the associated editing tools as well as performance issues. It is indeed important to keep *real-time performance* (cf. Axis 2)

throughout the interaction loop, from user inputs to display, via editing and rendering operations. Finally, it would be interesting to add *semantic information* into 2D or 3D geometric representations. Semantic geometry appears to be particularly useful for many applications such as the design of more efficient manipulation and animation tools, for automatic simplification and abstraction, or even for automatic indexing and searching. This constitutes a complementary but longer term research direction.

In the *MANAO* project, we want to investigate representations beyond the classical light, shape, and matter decomposition. We thus want to directly control the appearance of objects both in 2D and 3D applications (e.g., [98]): this is a core topic of computer graphics. When working with 2D vector graphics, digital artists must carefully set up color gradients and textures: examples range from the creation of 2D logos to the photo-realistic imitation of object materials. Classic vector primitives quickly become impractical for creating illusions of complex materials and illuminations, and as a result an increasing amount of time and skill is required. This is only for still images. For animations, vector graphics are only used to create legible appearances composed of simple lines and color gradients. There is thus a need for more complex primitives that are able to accommodate complex reflection or texture patterns, while keeping the ease of use of vector graphics. For instance, instead of drawing color gradients directly, it is more advantageous to draw flow lines that represent local surface concavities and convexities. Going through such an intermediate structure then allows to deform simple material gradients and textures in a coherent way (see Figure 8), and animate them all at once. The manipulation of 3D object materials also raises important issues. Most existing material models are tailored to faithfully reproduce physical behaviors, not to be *easily controllable* by artists. Therefore artists learn to tweak model parameters to satisfy the needs of a particular shading appearance, which can quickly become cumbersome as the complexity of a 3D scene increases. We believe that an alternative approach is required, whereby material appearance of an object in a typical lighting environment is directly input (e.g., painted or drawn), and adapted to match a plausible material behavior. This way, artists will be able to create their own appearance (e.g., by using our shading primitives [98]), and replicate it to novel illumination environments and 3D models. For this purpose, we will rely on the decompositions and tools issued from Axis 1.



(a) (b) (c) (d) (e) (f)

Figure 8. Based on our analysis [102] (Axis 1), we have designed a system that mimics texture (left) and shading (right) effects using image processing alone. It takes depth (a) and normal (d) images as input, and uses them to deform images (b-e) in ways that closely approximate surface flows (c-f). It provides a convincing, yet artistically controllable illusion of 3D shape conveyed through texture or shading cues.

POTIOC Team

3. Research Program

3.1. Introduction

The design of new user interfaces is a complex process that requires tackling research challenges at different levels. First, at a technological level, the input and output interaction space is becoming richer and richer. We will explore the new input/output modalities offered by such a technological evolution, and we will contribute to extend these modalities for the purpose of our main objective, which is to make 3D digital worlds available to all. Then, we will concentrate on the design of good interaction techniques that rely on such input/output modalities, and that are dedicated to the population targeted by this project, i.e. general public, specialists which are not 3D experts, and people with impairments. Finally, a large part of our work will be dedicated to the understanding and the assessment of user interaction. In particular, we will conduct user studies to guide the design of hardware and software UI, to evaluate them, and to better understand how a user interacts with 3D environments.

These three levels, input/output modalities, interaction techniques, and human factors will be the three main research directions of Potioc. Of course, they are extremely linked, and they cannot be studied independently, one after the other. In particular, user studies will follow the design process of hardware/software user interfaces from the beginning to the end, and both hardware and software exploration will be interdependent. The design of a new 3D user interface will thus require some work at different levels, as illustrated in Figure 2 . All members of Potioc will contribute in each of these research directions.

3.2. Exploring and enhancing input/output interaction space

The Potioc project-team will be widely oriented towards new innovative input and output modalities, even though standard approaches based on keyboard/mouse and standard screens will not be excluded. This includes motor-based interfaces, and physiological interfaces like BCI, as well as stereoscopic display and augmented reality setups. These technologies may have a great potential for opening 3D digital worlds to everyone, if they are correctly exploited.

We will explore various input/output modalities. Of course, we will not explore all of them at the same time, but we do not want to set an agenda either, for focusing on one of them. For a given need fed by end-users, we will choose among the various input/output modalities the ones that have the biggest potential. In the following paragraphs, we explain in more details the research challenges we will focus on to benefit from the existing and upcoming technologies.

3.2.1. Real-time acquisition and signal processing

There is a wide number of sensors that can detect users' activity. Beyond the mouse that detects x and y movements in the plane, various sensors are dedicated to the detection of 3D movements, pressure, brain and physiological activity, and so on. These sensors provide information that may be very rich, either to detect command intent from the user, or to estimate and understand the user's state in real-time, but that is difficutly exploitable as it. Hence, a major challenge here is to extract the relevant information from the noisy raw data provided by the sensor.

An example, and important research topic in Potioc, is the analysis of brain signals for the design of BCI. Indeed, brain signals are usually measured by EEG, such EEG signals being very noisy, complex and non-stationary. Moreover, for BCI-based applications, they need to be processed and analyzed in real-time. Finally, EEG signals exhibit large inter-user differences and there are usually few examples of EEG signals available to tune the BCI to a given user (we cannot ask the user to perform thousands of time the same mental task just to collect examples). As such, appropriate signal processing algorithms must be designed in order to robustly

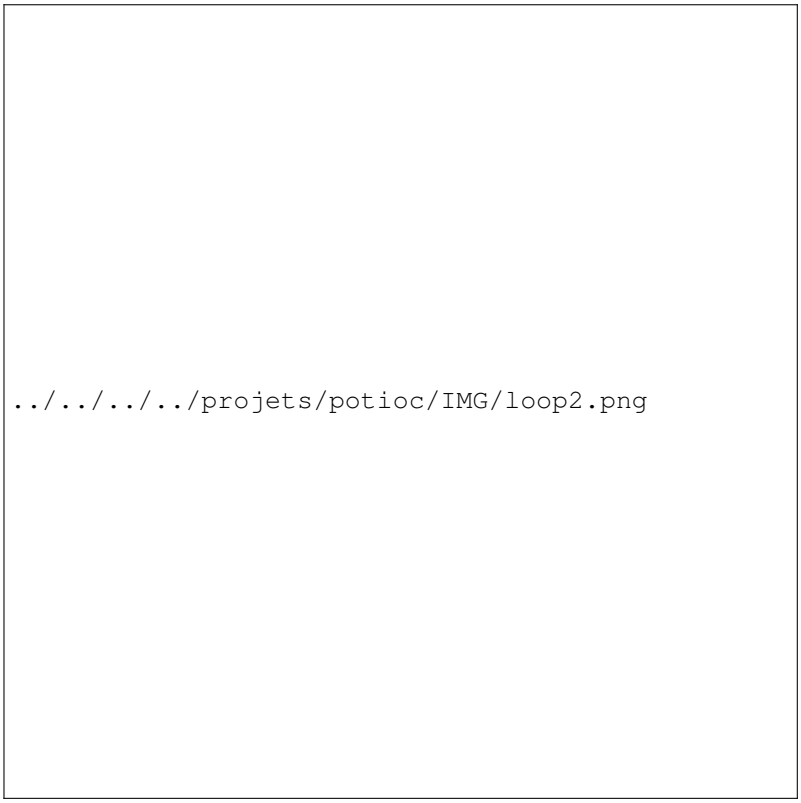


Figure 2. Diagram of an interactive system and the three main research axes of the Potioc project (blue boxes).

identify EEG patterns reflecting the user's intention. The research challenges are thus to design algorithms with high performances (in terms of rate of correctly identified user's state) anytime, anywhere, that are fully automatic and with minimal or no calibration time. In other words, we must design BCI that are convenient, comfortable and efficient enough so that they can be accepted and used by the end-user. Indeed, most users, in particular healthy users in the general public are used to highly convenient and efficient input devices (e.g., a simple mouse) and would not easily tolerate systems with a lower performance. Achieving this would make BCI good enough to be usable outside laboratories, e.g., for video gamers or patients. This will also make BCI valuable and reliable evaluation tools, e.g., to understand users' state during a given task. To address these challenges, pattern recognition and machine learning techniques are often used in order to find the optimal signal processing parameters. Similar approaches may contribute to the analysis of signals coming from other input devices than BCI. An example is the exploitation of depth cameras, where we need to find relevant information from noisy signals. Other emerging technologies will require similar attention, where the goal will be to transform an unstructured raw signal into a set of higher level descriptors that can be used as input parameters for controlling interaction techniques.

3.2.2. Restitution and perceptive feedback

Similarly to the input side, the feedback provided to the user through various output modalities will be explored in Potioc. Beyond the standard screens that are commonly used, we will explore various displays. In particular, in the scope of visual restitution, we will notably focus on large screens and tables, mobile setups and projection on real objects, and stereoscopic visualization. The challenge here will be to conceive good visual metaphors dedicated to these unconventional output devices in order to maximize the attractiveness and the pleasure linked to the use of these technologies.

For example, we will investigate the use of stereoscopic displays for extending the current visualization approaches. Indeed, stereoscopic visualization has been little explored outside complex VR setups dedicated to professional users and 3DTV. We believe that this modality may be very interesting for non-expert users, in wider contexts. To reach this goal, we will thus concentrate on new visual metaphors that benefit from stereoscopic visualization, and we will explore how, when, and where stereoscopy may be used.

Depending on the targeted interaction tasks, we may also investigate various additional output modalities such as tangible interaction, audio displays, and so on. In any case, our approach will be the same: understanding how new perceptive modalities may push the frontier of our current interactive systems.

3.2.3. Creation of new systems

In addition to the exploration and the exploitation of existing input and output modalities for enhancing interaction with 3D content, we may also contribute to extend the current input/output interaction space by building new interactive systems. This will be done by combining hardware components, or by collaborating with mechanics/electronics specialists.

3.3. Designing targeted interaction techniques

In the previous section, we focused on the input/output interaction space, which is closely related to hardware components. In this part, we focus on the design of interaction techniques, which we define here as the means through which a user will complete an interaction task in a given interaction space. Even if this is naturally also linked to the underlying hardware components, the research conducted in this axis of the project will mainly concern software developments.

Similar to the input/output interaction space, the design of interaction techniques requires focusing on both the motor and the sensory components. Thus, in our 3D spatial context, the challenges will be to find good mappings between the available input and the DOF that need to be controlled in the 3D environment, and to provide relevant feedback to users so that they can understand well what they are doing.

The design of interaction techniques should be strongly guided by the targeted end-users. For example, a 3D UI dedicated to an expert user will not suit a novice user, and the converse is also true. In Potioc, where the final goal is to open 3D digital worlds to anyone, we will concentrate on the general public, specialists that are not 3D experts, and people with impairments.

3.3.1. General public

3D UIs have mainly been designed for professional use. For example, modeling tools require expertize to be used correctly and, consequently, they exclude the general public from the process of creating 3D content. Similarly, immersive technologies have been dedicated to professional users for a long time. Therefore, immersive 3D interaction techniques have generally been thought for trained users, and they may not fit well with a general public context. In Potioc, an important motivation will be to re-invent 3D UIs to adapt them to the general public. This motivation will guide us towards new approaches that have been little explored until now. In particular, to reach our objective, we will give a strong importance to the following criteria:

- Intuitiveness: a very short learning curve is required.
- Enjoyability: this is needed to motivate novice users in the complex process of interaction with 3D content.
- Robustness: the UIs should support untrained users that may potentially interact with unpredictable actions.

In addition, we will keep connected with societal and technological factors surrounding the general public. For example, [multi]touch-screens have become very popular these past few years, and everyone tends to be familiar with a standard gesture vocabulary (e.g. pinch gestures and flicking gestures). We will rely on these commonly acquired *ways-of-interact* to optimize the acceptability of the 3D UIs we will design. In this part of the project the challenge will be to conceive 3D UIs that offer a high degree of interactivity, while ensuring an easy access to technology, as well as a wide adherence.

3.3.2. Specialists

General public will be one of the main targets of Potioc for the design of 3D UIs. However, we do not exclude specialists, who have little experience with 3D interaction. These specialists can be for example artists, archaeologists, or architects. In any case, we are convinced that 3D digital worlds could benefit to such categories of users if we propose dedicated 3D UIs that allow them to better understand, communicate, or create, with their respective skills. Because such specialists will gain expertize while interacting with 3D content, it will be necessary to design 3D UIs that can adapt to their evolving level of expertize. In particular, the UIs should be easy to use and attractive enough to encourage new users. At the same time, they should provide advanced features that the specialist can discover while gaining expertize.

3.3.3. People with impairments

While the general public has been only scarcely considered as a potential target audience for 3D digital worlds, another category of users is even more neglected: people with impairments. Indeed, such people, in particular those with motor impairments, are unable to use classical input devices, since they have been designed for healthy users. People with motor impairment have to use dedicated input devices, adapted to their disabilities, such as a single switch. Since such input devices usually have much fewer degrees of freedom than classical devices, it is necessary to come up with appropriate interaction techniques in order to efficiently use this limited number of DOF to still enable the user to perform complex tasks in the 3D environment. In Potioc, our focus will be on the use of BCI to enable motor impaired users to interact with 3D environment for learning, creation and entertainment. Indeed, BCI enable a user to interact without any motor movement.

3.4. Understanding and assessing user interaction

The exploration of the input/output interaction space, and the design of new interaction techniques, are strongly linked with human factors, which will be the third research axis of the Potioc project. Indeed, to guide the developments described in the previous sections, we first need to well understand users' motor and cognitive skills for the completion of 3D interaction tasks. This will be explored thanks to *a-priori* experiments. In order to evaluate our hardware and software interfaces, we will conduct *a-posteriori* user studies. Finally, we will explore new approaches for a real-time cognitive analysis of the performance and the experience of a user interacting with a 3D environment.

The main challenge in this part of the project will be to design good experimental protocols that will allow us to finely analyze various parameters for improving our interfaces. In 2D, there exist many standard protocols and prediction laws for evaluating UIs (e.g. Fitts law and ISO 9241). This is not the case in 3D. Consequently, a special care must be taken when evaluating interaction in 3D spatial contexts.

In addition to the standard experiments we will conduct in our lab, we will conduct large scale experiments thanks to the strong collaboration we have with the center for the widespread diffusion of scientific culture, Cap Sciences (see Collaboration section). With such kind of experiments, we will be able to test hundreds of participants of various ages, gender, or level of expertize that we will be able to track thanks to the Navinum system [1], and this during long a period of time. A challenge for us will be to gain benefit from this wealth of information for the development of our 3D UIs.

3.4.1. A-priori user studies

Before designing 3D UIs, it is important to understand what a user is good at, and what may cause difficulties. This is true at a motor level, as well as a cognitive level. For example, are users able to coordinate the movements of several fingers on a touchscreen at the same time, or are they able to finely control the quantity of force applied on it while moving their hand? Similarly, are the users able to mentally predict a 3D rotation, and how many levels of depth are they able to distinguish when visualizing stereoscopic images? To answer these questions, we will conduct preliminary studies.

Our research in that direction will guide our developments for the other research axes described above. For example, it will be interesting to explore touch-based 3D UIs that take into account several levels of force if we see that this parameter can be easily handled by users. On the other hand, if the results of a-priori tests show that this input cannot be easily controlled, then we will not push forward that direction.

The members of Potioc have already conducted such kinds of experiments, and we will continue our work in that direction. For some investigations, we will collaborate with psychologists and experts in cognitive science (see Collaborations section) to explore in more depth motor and cognitive human skills.

A-priori studies will allow us to understand how users tend to "*naturally*" interact to complete 3D interaction tasks, and to understand which feedbacks are the best suited. This will be a first answer to our global quest of providing pleasant interfaces. Indeed, this will allow us to adapt the UIs to the users, and not the opposite. This should enhance the global acceptability and motivation of users facing a new interactive system.

3.4.2. A-posteriori user studies

In Potioc, we will conceive new hardware and software interfaces. To validate these UIs, and to improve them, we will conduct user experiments, as classically done in the field of HCI. This is a standard methodology that we currently follow (see Bibliography). We will do this in our lab, and in Cap Sciences.

Beyond the standard evaluation criteria that are based on performance for speed, accuracy, coordination, and so on, we will also consider other criteria that are more relevant for the Potioc project. Indeed, we will give a great importance to enjoyability, pleasure of use, accessibility, and so on. Consequently, we will need to redefine the standard way to evaluate UIs. Once again, our relationship with Cap Sciences will help us in such investigations. The use of questionnaires will be a way to better understand how an interface should be designed to reach a successful use. In addition, we will observe and analyze how visitors tend to interact with various interfaces we will propose. For example, we will collect information like the time spent on a given interactive system or the number of smiles recorded during an interaction process. The identification of good criteria to use for the evaluation of a popular 3D UI will be one of the research directions of our team.

Conducting such *a-posteriori* studies, in particular with experts of mediation, with new criteria of success, will be a second answer to our goal of evaluating the pleasure linked to the use of 3D UIs.

[1] Navinum is a system based on a RFID technology that is used to collect informations about the activity of the visitors in Cap Sciences. http://www.scribd.com/doc/55178878/Dossier-de-Presse-Numerique-100511

3.4.3. Real-time cognitive analysis

Classically, the user's subjective preferences for a given 3D UI are assessed using questionnaires. While these questionnaires provide important information, this is only a partial, biased, a-posteriori/a-priori measure, since they are collected before or after the 3D interaction process. When questionnaires are administered during 3D interaction, this interrupts and disturbs the user, hence biasing the evaluation. Moreover, while evaluating performance and usefulness is now well described and understood, evaluating the user's experience and thus the system usability appears as much more difficult, with a lack of systematic and standard approaches. Ideally, we would like to measure the user response and subjective experience while he/she is using the 3D UI, i.e., in real-time and without interrupting him/her, in order to precisely identify the UI pros and cons. Questionnaires cannot provide such a measure.

Fortunately, it has been recently shown that BCI could be used in a passive way, to monitor the user's mental state. More precisely, recent results suggested that appropriately processed EEG signals could provide information about mental states such as error perception, attention or mental workload. As such, BCI are emerging as a new tool to monitor a user's mental state and brain responses to various stimuli, in real-time. In the Potioc project, we propose a completely new way to evaluate 3DUI: rather than relying only on questionnaires to estimate the user's subjective experience , we propose to exploit passive BCI to estimate the user's mental state in real-time, without interrupting nor disturbing him or her, while he/she is using the 3DUI. In particular, we aim at measuring and processing EEG and other biosignals (e.g., pulse, galvanic skin response, electromyogram) in real-time in order to estimate mental states such as interaction error potentials or workload/attention levels, among others. This will be used to finely identify how intuitive, easy-to-use and (ideally) enjoyable any given 3D UI is. More specifically, it will allow us to identify how, when and where the UI has flaws. Because the analysis will occur in real-time, we will potentially be able to modify the interface while the user is interacting. This should lead to a better understanding of 3D interaction. The work that will be achieved in this area could potentially also be useful for 2D interface design. However, since Potioc's main target is 3DUI, we will naturally focus the real-time cognitive evaluations on 3D contexts, with specific targets such as depth perception, or perception of 3D rotations.

This real-time cognitive analysis will be a third answer to reach the objectives of Potioc, which are to open 3D digital worlds to everyone by increasing the pleasure of use.