Activity Report 2012

Section Application Domains

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

4. Application Domains

4.1. Number theory

Being able to compute quickly and reliably algebraic invariants is an invaluable aid to mathematicians: It fosters new conjectures, and often shoots down the too optimistic ones. Moreover, a large body of theoretical results in algebraic number theory has an asymptotic nature and only applies for large enough inputs; mechanised computations (preferably producing independently verifiable certificates) are often necessary to finish proofs.

For instance, many Diophantine problems reduce to a set of Thue equations of the form $P(x, y) = a$ for an irreducible, homogeneous $P \in \mathbb{Z}[x, y]$, $a \in \mathbb{Z}$, in unknown integers x, y. In principle, there is an algorithm to solve the latter, provided the class group and units of a rupture field of P are known. Since there is no other way to prove that the full set of solutions is obtained, these algebraic invariants must be computed and certified, preferably without using the GRH.

Deeper invariants such as the Euclidean spectrum are related to more theoretical concerns, e.g., determining new examples of principal, but not norm-Euclidean number fields, but could also yield practical new algorithms: Even if a number field has class number larger than 1 (in particular, it is not norm-Euclidean), knowing the upper part of the spectrum should give a *partial* gcd algorithm, succeeding for almost all pairs of elements of $\mathcal{O}_K$. As a matter of fact, every number field which is not a complex multiplication field and whose unit group has rank strictly greater than 1 is almost norm-Euclidean [37], [38].

Algorithms developed by the team are implemented in the free PARI/GP system for number theory maintained by K. Belabas, which is a reference and the tool of choice for the worldwide number theory community.

4.2. Cryptology

Public key cryptology has become a major application domain for algorithmic number theory. This is already true for the ubiquitous RSA system, but even more so for cryptosystems relying on the discrete logarithm problem in algebraic curves over finite fields [7]. For the same level of security, the latter require smaller key lengths than RSA, which results in a gain of bandwidth and (depending on the precise application) processing time. Especially in environments that are constrained with respect to space and computing power such as smart cards and embedded devices, algebraic curve cryptography has become the technology of choice. Most of the research topics of the LFANT team concern directly problems relevant for curve-based cryptology: The difficulty of the discrete logarithm problem in algebraic curves determines the security of the corresponding cryptosystems. Complex multiplication, point counting and isogenies provide, on one hand, the tools needed to create secure instances of curves. On the other hand, isogenies have been found to have direct cryptographic applications to hash functions [39] and encryption [43]. Pairings in algebraic curves have proved to be a rich source for novel cryptographic primitives. Class groups of number fields also enter the game as candidates for algebraic groups in which cryptosystems can be implemented. However, breaking these systems by computing discrete logarithms has proved to be easier than in algebraic curves; we intend to pursue this cryptanalytic strand of research.

Apart from solving specific problems related to cryptology, number theoretic expertise is vital to provide cryptologic advice to industrial partners in joint projects. It is to be expected that continuing pervasiveness and ubiquity of very low power computing devices will render the need for algebraic curve cryptography more pressing in coming years.

ALEA Project-Team

4. Application Domains

4.1. Application Domains

This short section is only concerned with the list of concrete application domains developed by our team project on Bayesian inference and unsupervised learning, nonlinear filtering and rare event analysis. Most of these application areas result from fruitful collaborations with other national institutes or industries.

Three application domains are directly related to evolutionary computing, particle filtering and Bayesian inference. They are currently investigated by our team project:

1. **Multi-object tracking**. Multi-object tracking deals with the task of estimating the states of a set of moving objects from a set of measurements obtained sequentially. These measurements may either arise from one of the targets or from clutter and the measurement-to-target association is generally unknown. This problem can then be recast as a dynamic clustering one where the clusters are the clutter and the different targets. The targets actually move in time, some targets may appear/disappear over time and the number of targets is generally unknown and time-varying. The ALEA team has been involved in the ANR project Propagation (2010-2012) with DCNS, Thalés and Exavision.

2. **Finance**. The Team ALEA is interested in the design and analysis of new advanced particle methods for option pricing, partial observation problems, and sensitivity measures computation. An international workshop has been jointly organized by ALEA and CMAP (Polytechnique) on this topic in October 2012 (organizers: E. Gobet, P. Del Moral, P. Hu).

3. **Epidemiology**. Our team is interested in the development and analysis of particle mean field models for the calibration and uncertainty propagations in complex kinetic population models. The ALEA team is involved in an interdisciplinary exploratory research project with Laboratory Ecologie & Evolution, and co-organized an international workshop on this topic in 2011.

4. Application Domains

4.1. Introduction

We are working on problems that can be written in the following form

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F}_e(\mathbf{U}) - \nabla \cdot \mathbf{F}_v(\mathbf{U}, \nabla \mathbf{U}) = 0 \tag{1}$$

in a domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, subjected to initial and boundary conditions. The variable $\mathbf{U}$ is a vector in general, the flux $\mathbf{F}_e$ is a tensor, as well as $\mathbf{F}_v$ which also depends on the gradient of $\mathbf{U}$. The subsystem

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F}_e(\mathbf{U}) = 0$$

is assumed to be hyperbolic, the subsystem

$$\frac{\partial \mathbf{U}}{\partial t} - \nabla \cdot \mathbf{F}_v(\mathbf{U}, \nabla \mathbf{U}) = 0$$

is assumed to be elliptic. Last, (1) is supposed to satisfy an entropy inequality. The coefficients or models that define the flux and the boundary conditions can be deterministic or random.

The systems (1) are discretised mesh made of conformal elements. The tessalation is denoted by $\mathcal{T}_h$. The simplicies are denoted by K_j, $j = 1, n_e$, and $\cup_j K_j = \Omega_h$, an approximation of Ω. The mesh is assumed to be adapted to the boundary conditions. In our methods, we assume a globaly continuous approximation of $\mathbf{U}$ such that $\mathbf{U}_{|K_j}$ is either a polynomial of degree k or a more complex approximation such as a Nurbs. For now k is uniform over the mesh, and let us denote by V_h the vector space spanned by these functions, taking into account the boundary conditions.

The schemes we are working on have a variational formulation: find $\mathbf{U} \in V_h$ such that for any $\mathbf{V} \in V_h$,

$$a(\mathbf{U}, \mathbf{V}; \mathbf{U}) = 0.$$

The variational operator $a(\mathbf{U}, \mathbf{V}; \mathbf{W})$ is a sum of local operator that use onlty data within elements and boundary elements: it is very local. Boundary conditions can be implemented in a variational formulation of via penalisation, see figure 3 . The third argument $\mathbf{W}$ stands for the way are implemented the non oscillatory properties of the method.

This leads to highly non linear systems to solve, we use typicaly non linear Krylov space techniques. The cost is reduced thanks to a parallel implementation, the domain is partionnned via `Scotch`. Mesh balancing, after mesh refinement, is handled via `PaMPA`. These schemes are implemented in `RealfluiDS` and, partialy, `Aerosol`. An example of such a simulation is given by Figure 2 .

In case of non demistic problems, we have a semi-intrusive strategy. The randomness is expressed via N scalar random parameters (that might be correlated), $X = (x_1, ..., x_N)$ with probability measure $d\mu$ which support is in a subset of $\mathbb{R}^N$. The idea of non intrusive methods is to approximate $d\mu$ either by $d\mu \approx \sum_j \omega_l \delta_{X_l}$ for $\omega_l \geq 0$ that sum up to unity, for "well chosen" samples X_l or by $d\mu \approx \sum_l \mu(\Omega_l) 1_{X_j} dX$ where the sets Ω_j covers the support of $d\mu$ and are non overlapping.
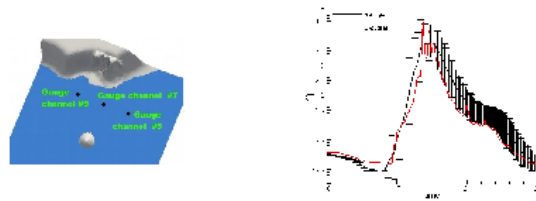
Figure 1. Topography for the Okushiri tsunami with random amplitude and phase. The mean and variance of the average at gauge 9.
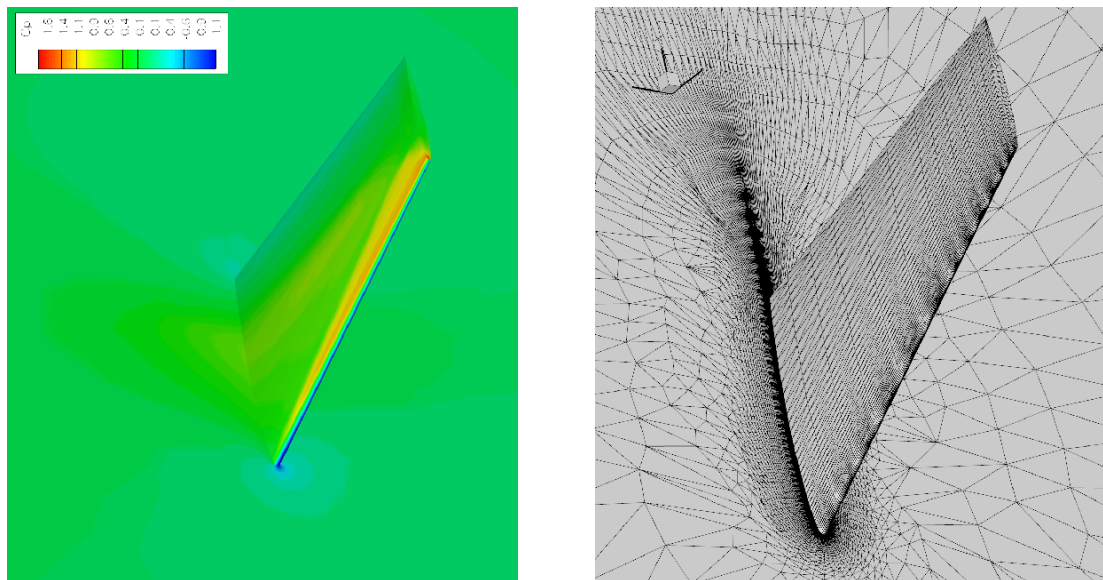

Figure 2. Flow over a turbulent M6 wing (pressure coefficient, mesh by Dassault Aviation).
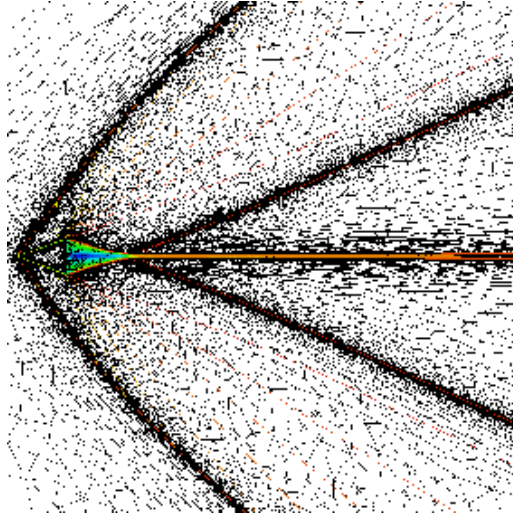
Figure 3. Adapted mesh for a viscous flow over a triangular wedge.

Staring from a discrete approximation of (1), we can implement randomess in the scheme. An example is given on figure 1 applied to the shallow water equations with dry shores, when the amplitude of the incoming tsunami wave is not known.

4.2. External and internal Aerodynamics

A classical application is the simulation of internal and external flows, with perfect or real gas equation of states, in complex geometries. This requires often the use of meshes having heterogeneous structures. We are working with unstructured meshes, either with simplicial elements or mixtures of hex, tets, pyramids and prisms. Mesh refinement can be enable in order to better resolve either the discontinuous flow structures or the capture of boundary layers.

4.3. Multiphase flows

Another domain of application is the simulation of multiphase flows. Here, the system (1) need to be supplemented by at least a PDE describing the phase volume changes, and the equation of states of the phases. The system is in most case writetn in non conservative form, so that additional difficulies need to be handled.

Multiphase fows occur in many applications: for petroleum industry, nuclear industry (accident management), engines, pipes, etc.

4.4. Inflight icing and ice shedding

Actual concerns about greenhouse gases lead to changes in the design of aircraft with an increase use of composite materials. This in turns offers new possibilities for design of ice protection systems, thus renewing interest in de-icing simulation tools. To save fuel burn, aircraft manufacturers are investigating ice protection systems such as electro-thermal or electro-mechanical de-icing systems to replace anti-icing systems. By reducing the adhesive shear strength between ice and surface, de-icing systems remove ice formed on the protected surfaces following a periodic cycle. This cycle is defined such that inter cycle ice shapes remain acceptable from a performance point of view. One of the drawbacks of de-icing device is the ice pieces shed into the flow. The knowledge of ice shedding trajectories could allow assessing the risk of impact/ingestion

on/in aircraft components located downstream. When the pieces leave the aircraft surface, they become projectiles that can hit and cause severe damage to aircraft surface or other components, such as aircraft horizontal and vertical tails, or aircraft engine. Aircraft certification authorities, such as FAA, have specific requirements for large ice fragment ingestion during engine certification. Control surfaces or wing flaps are also sensitive to ice shedding because they can be blocked by ice fragments. Aircraft manufacturers rely mainly on flight tests to evaluate the potential negative effects of ice shedding because of the lack of appropriate numerical tools. The random shape and size taken by ice shed particles together with their rotation as they move make it difficult for classical CFD tools to predict trajectories. The numerical simulation of a full unsteady viscous flow, with a set of moving bodies immersed within, shows several difficulties for grid based methods. Drawbacks income from the meshing procedure for complex geometries and the re-griding procedure in tracing the body motion. A new approach that take into account the effect of ice accretion on flow field is used to solve the ice trajectory problem. The approach is based on mesh adaptation, penalization method and level sets.

4.5. Modeling study of Silicic acid flux to the ocean from tidal permeable sediments (in collaboration with EPOC)

Sandy sediments of tidal beaches are poor in reactive substances because they are regularly flushed by significant flow caused by tidal forcing.This transport process may significantly affect the flux of reactive solutes to the ocean. A two dimensional model coupling the Richards equation that describes the flow in permeable sediments and the conservation equation of the silicic acid was developed to simulate the evolution of the silicic acid concentration into a variably saturated porous media submitted to tidal forcing.

4.6. ORCs cycles

ORCs are Rankine cycles that use properly chosen low-boiling molecularly heavy organic compounds to drive the turbine in place of steam. This makes them suitable for the exploitation of low grade heat sources like biomass combustion, geothermal reservoirs and heat recovery from industrial processes. ORC turbines mainly use a single (less frequently, two) stage to expand the fluid. Up till present, no experimental data are available for flows of heavy fluids in the dense gas region. Experiments are difficult be cause of high temperature and pressure conditions, and fluid decomposition or inflammability in presence of air. This has motivated the use of numerical simulation as a preferential tool for dense gas flow analysis, but only a limited number of papers have been devoted to the computation of dense gas flows. With no experimental validation yet available for any of these configurations, care must be taken in the analysis of the computed flow fields because of their sensitivity to the thermodynamic model and to the numerical ingredients of the discretization scheme. Since no comparison with experimental data is possible, particular attention is devoted to code validation and model assessment. We created the plateform ORComp, for computing some global performance metrics, and we applied some UQ and numerical methods for taking into account the sun variability in the design of ORCs cycles.

4.7. Cavitation

Cavitation consists in a local pressure drop below the vapor pressure at the liquid temperature, thus creating a phase change and vapor bubbles formation. Their collapse in high-pressure region can dramatically lead to failure, erosion and other undesirable effects. For this reason, there is a strong effort devoted to develop predictive numerical tools for cavitating flows in industrial applications. Unfortunately, an accurate description of interactions between the vapour and liquid phases requires accurate physical models and a way to take into account the dynamics of the interface. Moreover, multiscale effects, turbulence and thermodynamics should be also considered. Cavitation models are typically dependent on two types of parameters: first, on some physical parameters, such as for example the number of bubbles, that is not usually well measured; secondly, on some empiric parameters, useful for fitting and calibration procedures with respect to the experimental data. Therefore, model parameters represent an impor- tant source of uncertainty. Moreover, it is not an easy task to

well define boundary and initial conditions, because of difficulties encountered in order to control accurately experiments in cavitating flows. As a result, conditions imposed for the setting of a numerical simulation, are affected by a dramatic randomness. We performed a systematic study for considering the probabilistic properties of the input parameters permitting to capture non-linearities in uncertainty propagation. Moreover, the DEM method has been modified to take into account cavitation phenomena with real-gas effects.

4.8. Atmospheric entries of spacecraft

Simulation of atmospheric entries of spacecraft is a challenging problem involving many complex physical phenomena, including rarefied gas effects, aerothermochemistry, radiation, and the response of thermal protection materials to extreme conditions. The post-flight analysis of a space mission requires accurate determination of the freestream conditions for the trajectory, that is, temperature and pressure conditions and the Mach number in front of the shock. The latters can be rebuilt from the pressure and heat flux measured on the spacecraft by means of a Flush Air Data System (FADS). This instrumentation comprises a set of sensors flush mounted in the thermal protection system to measure the static pressure (pressure taps) and heat flux (calorimeters). In this context, Computational Fluid Dynamics (CFD) supplied with UQ tools permits to take into account chemical effects and to include both measurement errors and epistemic uncertainties on the chemical model parameters in the bulk and at the wall (surface catalysis). Rebuilding the freestream conditions from the FADS data therefore amounts to solving a stochastic inverse problem. In this contexte, we proposed a new methodology for solving the inverse problem based on a Bayesian setting, that is, probability densities of possible values of freestream conditions are rebuilt from stagnation-point pressure and heat flux measurements. A Bayesian setting offers a rigorous foundation for inferring input parameters from noisy data and uncertain forward models, a natural mechanism for incorporating prior information, and a quantitative assessment of uncertainty on the inferred results.

4.9. Shallow Water

Prediction of shallow water equations in realistic application depends on the level of complexity used for the physical modelling (such as for example, for the friction coefficient) and on a set of empirical coefficients that are usually chosen in order to fit the experimental data. Then, input environmental conditions, topography and modelling involve a certain degree of uncertainty. The capability to take into account these uncertainties in the numerical simulation is of great importance in order to correctly predict extreme flood events. Stochastic modeling of long-wave propagation demands a robust shallow-water model in order to characterize the physical processes. We coupled a residual distribution scheme for shallow water equations with some stochastic methods in order to take into account uncertainties in the numerical simulation. Preliminary results showed that influence of uncertainties is stronger after the phase interaction indicating the need for a stochastic simulation in order to have a correct prediction of the numerical solution.

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

4. Application Domains

4.1. Effusion cooling of aeronautical combustion chambers walls

The industrial applications of our project is the cooling of the walls of the combustion chambers encountered in the helicopter engines, and more precisely, we wish to contribute to the improvement of effusion cooling.

Effusion cooling is nowadays very widespread, especially in the aeronautical context. It consists in piercing holes on the wall of the combustion chamber. These holes induce cold jets that enter inside the combustion chamber. The goal of this jet is to form a film of air that will cool the walls of the chamber, see Figure 2 .
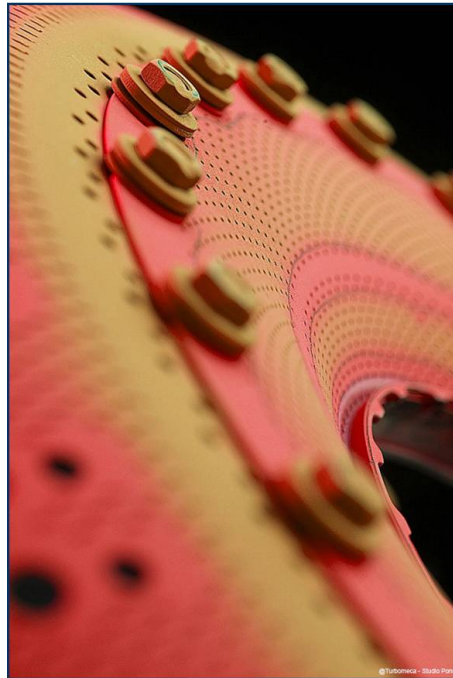

Figure 2. Effusion cooling of aeronautical combustion chambers: close view of a typical perforated chamber wall

Effusion cooling in a combustion chamber takes at the wall where thousands of small holes allow cool air to enter inside the combustion chamber. This induces jets in crossflow in charge of cooling the walls, whatever the heat and the acoustic waves present inside the chamber. Nevertheless, this technique is not straightforward to put in practice: the size, design and position of the holes can have an important effect on the cooling efficiency. For a safe and efficient functioning of the combustion chamber, it is required that the cooling jets and the combustion effects be as much independent as possible. For example, this means that

- The jets of cool air should not mix too much with the internal flow. Otherwise it will decrease the efficiency of the combustion.
- The jets should be as much stable as possible when submitted to waves emitted in the combustion chamber, e.g. acoustic waves induced by combustion instabilities. Otherwise the jets may not cool enough the walls of the combustion chamber which can then undergoes severe damages.

The first point is what we aim at simulate in this project. As the model chosen is the fully compressible Navier Stokes system, there should not be any problem in the future for being able to simulate the effect of an acoustic forcing on the jet in crossflow.

Having a database of Direct Numerical Simulations is also fundamental for testing closure laws that are used in turbulence models encountered in RANS and LES models. With such models, it is possible for example to perform optimisation.

A last aspect, that will not be dealt with in this project, but that could be dealt with in the future, is the interaction between the flow and the wall. The aim is to understand the effect of coupling between the heat propagation in the wall and the flow near the wall. A careful study of this interaction can allow to determine the exchange coefficients, and so the efficiency of the cooling by the jet. Such determination may particularly useful to develop one or multidimensional models of wall-fluid interaction [19].

From the application point of view, compressibility effects must be taken into account since the Mach number of the flow can reach values equal to 0.3, hence/or acoustic waves may be present inside the combustion chamber. This can raise a problem, because upwind numerical schemes are known to be less accurate in the low Mach limit.

CONCHA Project-Team

4. Application Domains

4.1. Aerodynamics

Aerodynamics provide a challenging field for numerical simulations in fluid dynamics with a wide range of applications. Robustness of the simulation software with respect to physical parameters as the Reynolds and Mach numbers is necessary condition. In general, realistic simulations need to be done in three dimensions, which makes the efficiency of the numerical approach and implementation a question of feasibility. Therefore, different efforts are made in this project in order to tackle these subjects.

4.2. Red blood cells

Hammou El-Otmany started his PhD thesis in October 2012 in our group, supervised by D. Capatina and D. Graebling. The thesis is financed by UPPA (50%) and CDAPP (50%) and concerns the numerical simulation of biological fluids flows. We will focus more particularly on the physical and numerical modeling of red blood cells.

Clinically, some pathologies such as drepanocytosis or sickle cell anemia are due to the abnormal form of red blood cells. In the microcirculation, where cells must deform to pass through narrow capillaries, the deformability of individual red blood cells is a major determinant of resistance to flow.

The goal is twofold. On the one hand, we want to propose a realistic modeling of red blood cells in artery flow, by taking into account the membrane's viscoelasticity and thus, its deformability. The latter is essentially linked to its structure (i.e. its cellular geometry, membrane properties and cytoplasmic viscosity); thus structural abnormalities, as found in some haematological disorders can be expected to affect blood flow in the microcirculation and/or red cell lifespan.

On the other hand, we want to develop an efficient and stable numerical method in order to treat the coupling between the different models involved: Navier-Stokes for the matrix (blood) and for the cytoplasme (interior of the cell) and a non-Newtonian fluid (for instance, Giesekus) for the membrane. We will use the NXFEM method to take into account the interfaces between fluids.

4.3. Heat transfer

Heat transfer problems involve the coupling of the flow field of the fluid with temperature inside the flow and possibly on the boundary of the flow domain. A typical example of a heat transfer problem is the cooling of a combustion engine, see the project Optimal described in Section 7.1 .

4.4. Turbulence

Turbulent flows are ubiquitous in industrial applications. Direct numerical simulation (DNS), which aims at complete resolution of the flow field up to the Kolmogorov scale, has historically been limited to very simple geometries. The increase of computational power and the development of specialized numerical methods open the door to a wider range of applications. However, for most applications of practical interest, the use of some kind of turbulence modeling is unavoidable in order to obtain the prediction of averaged values and commercial software is in general based on such approaches combined with wall laws. In many applications, such as the project Optimal, see Section 7.1 , the Reynolds number is at an intermediate level, which means that the turbulence is not fully developed, and the heuristics behind most turbulence models are questionable. Especially, in heat transfer problems, the usage of wall laws seems to considerably lower the accuracy of the predicted mean values. In order to improve the computation of such values, we are particularly interested in variational multiscale methods and its relations to stabilized finite element methods.

4.5. Flows in porous media

Flows in fractured pourous media are very important in petroleum engineering. They represent a good framework for the application of the tools developed in the CONCHA library such as the NXFEM method, goal oriented adaptivity, multiscale coupling of different models and multilevel solvers.

CQFD Project-Team

4. Application Domains

4.1. Dependability and safety

Our abilities in probability and statistics apply naturally to industry in particular in studies of dependability and safety.

An illustrative example which gathers all the topics of team is a collaboration started in May 2010 with Thales Optronique on the subject of *optimization of the maintenance of a digital camera equipped with HUMS* (Health Unit Monitoring Systems). This subject is very interesting for us because it combines many aspects of our project. Classification tools will be used to select significant variables as the first step in the modeling of a digital camera. The model will then be analysed and estimated in order to optimize the maintenance.

A second example concerns the optimization of the maintenance date for an aluminum metallic structure subject to corrosion. It is a structure of strategic ballistic missile that is stored in a nuclear submarine missile launcher in peace-time and inspected with a given periodicity. The requirement for security on this structure is very strong. The mechanical stress exerted on the structure depends on its thickness. It is thus crucial to control the evolution of the thickness of the structure over time, and to intervene before the break.

A third example is the minimization of the acoustic signature of a submarine. The submarine has to chose its trajectory in order to minimize at each time step its observability by a surface ship following an unknown random trajectory.

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

GEOSTAT Project-Team

4. Application Domains

4.1. Application Domains

In GEOSTAT, the development of nonlinear methods for the study of complex systems and signals is conducted on four broad types of complex signals:

- Ocean dynamics and ocean/atmosphere interactions: generation of high-resolution maps from cascading properties and the determination of optimal wavelets[9], geostrophic or non-geostrophic-complex oceanic dynamics, mixing phenomena.

- Speech signal (analysis, recognition, classification)[1].

- Optimal wavelets for phase reconstruction in adaptive optics[2].

- Heartbeat signals, in cooperation with IHU LIRYC and Professor M. Haissaguerre (INSERM EA 2668 Electrophysiology and Cardiac Stimulation)[5] [35].

4. Application Domains

4.1. Introduction

We now present our contribution to these above challenges concerning interface problem for complex fluids, direct simulations and analysis, flow control and optimization. From the technical point of view, many productions are common to the different parts of the project. For example, level-set methods, fast-marching procedure are used for shape optimization and for microfluidics, penalization methods are used for high Reynolds flows and for tumor growth. This leads to a strong politic of development of numerical modules.

4.2. Multi-fluid flows

- computation of bifluid flows : see the thesis of S. Tancogne ([83]) and P. Vigneaux ([86]). Stability of an interface, shape of droplets, formation of a jet. Study of the Plateau-Rayleigh instability. Behaviour of diphasic fluids evolving in square microchannels.
- mixing in micro-channel : see the thesis of J. Dambrine [67]. Passive mixing strategies involving boundary conditions. Enhanced oil recovery (study of mixing oil-water-polymer in a microchannel).
- emulsions and foam : see the thesis of S. Benito [52]. Applications in biology : behaviour of tissues, of tumor,....
- polymer nanotube conglomerate wire : it was the subject of a talk in the following conference "WCCM8-ECCOMAS2008" and of the talk [65].

4.3. Cancer modeling

- specific models : investigation of two particular cancer : gliomas (brain tumors), colorectal cancers lung and lever metastasis, brest cancer. This is one part of the PhD works of J.B. Lagaert and D. Lombardi.
- modelling of electrochemotherapy : see ARC C3MB (http://www.math.u-bordeaux1.fr/ArcC3MB/)
- parameter estimations with the help of low order models : see the PhD of J.B. Lagaert and D. Lombardi
- patient-specific simulations
- optimal shape design : the goal is to recover the vascularization of a model tumor from the knowledge of its shape evolution.

4.4. Newtonian fluid flows simulations and their analysis

- simulation of a synthetic or pulsed jet. This is an ongoing project with Renault and PSA inside a PREDIT project.
- vortex dynamics : see [69].
- simulation of compressible flows on cartesian grids : see the thesis of Gabriele Ottino's Thesis [79], who underwent his doctoral studies in conjunction in the MC2 team and at the Politecnico di Torino, and defended in April 2009. He had a grant of the French-Italian university.
- 3D turbulent flows through DESGRIVRE contract with AIRBUS. Thesis of C. Wervaecke [88]. The goal is to use Detached-Eddy Simulation to model turbulent flows around iced bodies.
- porous media : Numerical study of coupling between Richards and transport-diffusion equations in permeable sediment affected by tidal oscillation. See the thesis of R. Chassagne [63]

4.5. Flow control and shape optimization

- passive control : the idea is to put a porous interface between the solid body and the fluid. See the D. Depeyras thesis [68] and Yong-Liang Xiang [90] and CH Bruneau and Iraj Mortazavi) [55]. See also project [60] founded by the Euopean Community.

- active control : see the three PhD thesis: M. Buffoni, J. Weller [87], E. Lombardi and FFAST project funded by EU and iled by the University of Bristol and AIRBUS UK.

- shape optimization for turbo-machines : See [84].

- reduced order models : it consists in designing a non-linear observer that estimates the state of the flow field from a limited number of measurements in the field. The challenge is to reduce as much as possible the information required and to take it from the boundary. See J. Weller [87] and E. Lombardi.

- passive control of flows with porous media : see [57], [54], [53], [78], [58].

- inverse problems in imagery : see [62].

REALOPT Project-Team

4. Application Domains

4.1. Introduction

Our group has tackled applications in logistics, transportation and routing [57], [56], [53], [55], in production planning [70] and inventory control [53], [55], in network design and traffic routing [39], [46], [51], [75], [36], [47], [61], [66], in cutting and placement problems [60], [67], [68], [69], [71], and in scheduling [7], [62], [34].

4.2. Network Design and Routing Problems

We are actively working on problems arising in network topology design, implementing a survivability condition of the form "at least two paths link each pair of terminals". We have extended polyhedral approaches to problem variants with bounded length requirements and re-routing restrictions [46]. Associated to network design is the question of traffic routing in the network: one needs to check that the network capacity suffices to carry the demand for traffic. The assignment of traffic also implies the installation of specific hardware at transient or terminal nodes.

To accommodate the increase of traffic in telecommunication networks, today's optical networks use grooming and wavelength division multiplexing technologies. Packing multiple requests together in the same optical stream requires to convert the signal in the electrical domain at each aggregation of disaggregation of traffic at an origin, a destination or a bifurcation node. Traffic grooming and routing decisions along with wavelength assignments must be optimized to reduce opto-electronic system installation cost. We developed and compared several decomposition approaches [77], [76], [75] to deal with backbone optical network with relatively few nodes (around 20) but thousands of requests for which traditional multi-commodity network flow approaches are completely overwhelmed. We also studied the impact of imposing a restriction on the number of optical hops in any request route [74]. We also developed a branch-and-cut approach to a problem that consists in placing sensors on the links of a network for a minimum cost [51], [52].

We studied several time dependent formulations for the unit demand vehicle routing problem [41], [40]. We gave new bounding flow inequalities for a single commodity flow formulation of the problem. We described their impact by projecting them on some other sets of variables, such as variables issued of the Picard and Queyranne formulation or the natural set of design variables. Some inequalities obtained by projection are facet defining for the polytope associated with the problem. We are now running more numerical experiments in order to validate in practice the efficiency of our theoretical results.

We also worked on the p-median problem, applying the matching theory to develop an efficient algorithm in Y-free graphs and to provide a simple polyhedral characterization of the problem and therefore a simple linear formulation [65] simplifying results from Baiou and Barahona.

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

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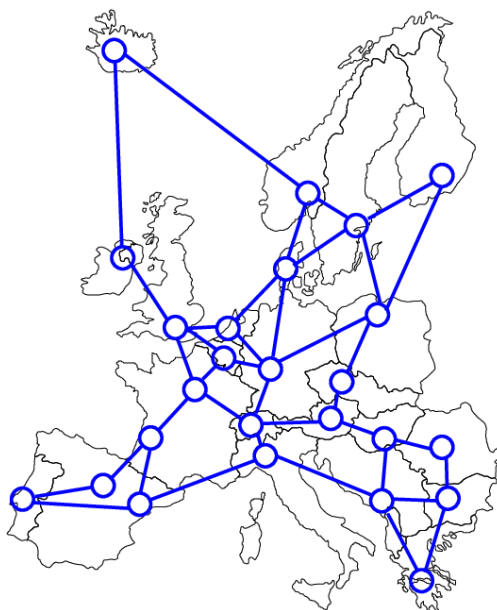
Figure 1. Design of a SDH/SONET european network where demands are multiplexed.

4.3. Packing and Covering Problems

We developed a branch-and-price algorithm for the Bin Packing Problem with Conflicts which improves on other approaches available in the literature [20]. The algorithm uses our methodological advances like the generic branching rule for the branch-and-price and the column based heuristic. One of the ingredients which contributes to the success of our method are fast algorithms we developed for solving the subproblem which is the Knapsack Problem with Conflicts. Two variants of the subproblem have been considered: with interval and arbitrary conflict graphs. The paper which presents this work is being finalized.

We have designed a new algorithm for vertex packing (equivalently stable set) in claw-free graphs [58]. Previously the best known algorithm for this problem had a running time of $O(n^6)$ (with n the number of vertices in the graph) while our new algorithm runs in $O(n^4)$.

We studied a variant of the knapsack problem encountered in inventory routing problem [55]: we faced a multiple-class integer knapsack problem with setups [54] (items are partitioned into classes whose use implies a setup cost and associated capacity consumption). We showed the extent to which classical results for the knapsack problem can be generalized to this variant with setups and we developed a specialized branch-and-bound algorithm.

We studied the orthogonal knapsack problem, with the help of graph theory [50], [48], [12], [11]. Fekete and Schepers proposed to model multi-dimensional orthogonal placement problems by using an efficient representation of all geometrically symmetric solutions by a so called *packing class* involving one *interval graph* for each dimension. Though Fekete & Schepers' framework is very efficient, we have however identified several weaknesses in their algorithms: the most obvious one is that they do not take advantage of the different possibilities to represent interval graphs. We propose to represent these graphs by matrices with consecutive ones on each row. We proposed a branch-and-bound algorithm for the 2d knapsack problem that uses our 2D packing feasibility check.

4.4. Planning, Scheduling, and Logistic Problems

Inventory routing problems combine the optimization of product deliveries (or pickups) with inventory control at customer sites. We considered an industrial application where one must construct the planning of single product pickups over time; each site accumulates stock at a deterministic rate; the stock is emptied on each visit. We have developed a truncated branch-and-price algorithm: periodic plans are generated for vehicles by solving a multiple choice knapsack subproblem; the global planning of customer visits is generated by solving a master program. Confronted with the issue of symmetry in time, we used a state-space relaxation idea. Our algorithm provides solutions with reasonable deviation from optimality for large scale problems (260 customer sites, 60 time periods, 10 vehicles) coming from industry [13]. We previously developed approximate solutions to a related problem combining vehicle routing and planning over a fixed time horizon (solving instances involving up to 6000 pick-ups and deliveries to plan over a twenty day time horizon with specific requirements on the frequency of visits to customers [57].

Together with our partner company GAPSO from the associate team SAMBA, we are working on the equipment routing task scheduling problem [28] arising during port operations. In this problem, a set of tasks needs to be performed using equipments of different types with the objective of maximizing the weighted sum of performed tasks.

We participated to the project on an airborne radar scheduling. For this problem, we developed fast heuristics [45] and exact algorithms [34]. A substantial research has been done on machine scheduling problems. A new compact MIP formulation was proposed for a large class of these problems [33]. An exact decomposition algorithm was developed for the NP-hard maximizing the weighted number of late jobs problem on a single machine [62]. A dominant class of schedules for malleable parallel jobs was discovered in the NP-hard problem to minimize the total weighted completion time [64]. We proved that a special case of the scheduling problem at cross docking terminals to minimize the storage cost is polynomially solvable [19], [63].

Another application area in which we have successfully developed MIP approaches is in the area of tactical production and supply chain planning. In [14], we proposed a simple heuristic for challenging multi-echelon problems that makes effective use of a standard MIP solver. [14] contains a detailed investigation of what makes solving the MIP formulations of such problems challenging; it provides a survey of the known methods for strengthening formulations for these applications, and it also pinpoints the specific substructure that seems to cause the bottleneck in solving these models. Finally, the results of [35] provide demonstrably stronger formulations for some problem classes than any previously proposed.

We have been developing **robust optimization** models and methods to deal with a number of applications like the above in which uncertainty is involved. In [43], [42], we analyzed fundamental MIP models that incorporate uncertainty and we have exploited the structure of the stochastic formulation of the problems in order to derive algorithms and strong formulations for these and related problems. These results appear to be the first of their kind for structured stochastic MIP models. In addition, we have engaged in successful research to apply concepts such as these to health care logistics [37]. We considered train timetabling problems and their re-optimization after a perturbation in the network [44]. The question of formulation is central. Models of the literature are not satisfactory: continuous time formulations have poor quality due to the presence of discrete decision (re-sequencing or re-routing); arc flow in time-space graph blow-up in size (they can only handle a single line timetabling problem). We have developed a discrete time formulation that strikes a compromise between these two previous models. Based on various time and network aggregation strategies, we develop a 2-stage approach, solving the contiguous time model having fixed the precedence based on a solution to the discrete time model.

CARMEN Team

4. Application Domains

4.1. Scientific context: the Liryc

Our fields of application are naturally: electrophysiology and cardiac physiopathology at the tissue scale on one side; medical and clinical cardiology on the other side.

The team's research project is part of the IHU Liryc project, initiated by Pr. M. Haissaguerre. It is concerned by the major issues of modern electrocardiology: atrial arrhythmias, sudden death due to ventricular fibrillation and heart failure related to ventricular dyssynchrony.

We aim at bringing applied mathematics and scientific computing closer to biomedical research applied to cardiac rhythmology and clinical cardiology. It aims at enhancing our fundamental knowledge of the normal and abnormal cardiac electrical activity, of the patterns of the electrocardiogram; and we will develop new simulation tools for training, biological and clinical applications.

4.2. Basic experimental electrophysiology

Our modeling is carried out in coordination with the experimental teams from the Liryc. It will help to write new concepts concerning the multiscale organisation of the cardiac action potentials and will serve our understanding in many electrical pathologies:

- At the atrial level, we apply our models to understand the mechanisms of complex arrythmias and the relation with the heterogeneities at the insertion of the pulmonary vein.

- At the ventricula level, we focus on (1) modeling the complex coupling between the Purkinje network and the ventricles and (2) modeling the structural heterogeneities at the cellular scale, taking into account the complex organisation and disorganisation of the myocytes and fibroblasts. Point (1) is supposed to play a major role in sudden cardiac death and point (2) is important in the study of infarct scars for instance.

4.3. Cardiac electrical signals

The Liryc use, on a daily basis and in the clinical context, complex electrical imaging systems, like intracardiac catheters and the CardioInsight vest with 252 body syrface electrodes.

The numerical models can guide the analysis of these signals and conversely, the models can be guided by the signals.

Other applied questions can be addressed by modeling, like the nature of the various electrical signals measured by catheters, that heavily depends on the nature and spatial localisation of the electrodes.

<p style="text-align:center;color:red;font-weight:bold;">MAGIQUE-3D Project-Team</p>

4. Application Domains

4.1. Seismic Imaging

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

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

MAGNOME Project-Team

4. Application Domains

4.1. Function and history of yeast genomes

Yeasts provide an ideal subject matter for the study of eukaryotic microorganisms. From an experimental standpoint, the yeast *Saccharomyces cerevisiae* is a model organism amenable to laboratory use and very widely exploited, resulting in an astonishing array of experimental results. From a genomic standpoint, yeasts from the hemiascomycete class provide a unique tool for studying eukaryotic genome evolution on a large scale. With their relatively small and compact genomes, yeasts offer a unique opportunity to explore eukaryotic genome evolution by comparative analysis of several species.

- Yeasts are widely used as cell factories, for the production of beer, wine and bread and more recently of various metabolic products such as vitamins, ethanol, citric acid, lipids, etc.

- Yeasts can assimilate hydrocarbons (genera *Candida*, *Yarrowia* and *Debaryomyces*), depolymerise tannin extracts (*Zygosaccharomyces rouxii*) and produce hormones and vaccines in industrial quantities through heterologous gene expression.

- Several yeast species are pathogenic for humans, especially *Candida albicans*, *Candida glabrata*, *Candida tropicalis* and the Basidiomycete *Cryptococcus neoformans*.

The hemiascomycetous yeasts represent a homogeneous phylogenetic group of eukaryotes with a relatively large diversity at the physiological and ecological levels. Comparative genomic studies within this group have proved very informative [33], [48], [47], [35], [50], [30], [31], [6].

MAGNOME applies its methods for comparative genomics and knowledge engineering to the yeasts through the ten-year old *Génolevures* program (GDR 2354 CNRS), devoted to large-scale comparisons of yeast genomes with the aim of addressing basic questions of molecular evolution. We developed the software tools used by the CNRS's `genolevures.org` web site. For example, MAGNOME's MAGUS system for simultaneous genome annotation combines semi-supervised classification and rule-based inference in a collaborative web-based system that explicitly uses comparative genomics to simultaneously analyse groups of related genomes.

4.2. Alternative fuels and bioconversion

Oleaginous yeasts are capable of synthesizing lipids from different substrates other than glucose, and current research is attempting to understand this conversions with the goal of optimizing their throughput, production and quality. From a genomic standpoint the objective is to characterize genes involved in the biosynthesis of precursor molecules which will be transformed into fuels, which are thus not derived from petroleum. Biological experimentation by partner laboratories study lipid accumulation the oleaginous yeasts such as *Yarrowia lipolytica* starting from:

- pentoses, produced from lignin cellulose agricultural substrates following a biorefining strategy,

- glycerol, a secondary output of chemical production of biodiesel, and

- industrial residues.

Lipases from *Y. lipolytica* are of particular interest (see [38] for review). Experimental characterization of the lipid bodies produced from these substrates will aid in the identification of target genes which may serve for genetic engineering. This in turn requires the development of molecular tools for this class of yeasts with strong industrial potential. MAGNOME's focus is in acquiring genome sequences, predicting genes using models learned from genome comparison and sequencing of cDNA transcripts, and comparative annotation. Our overall goal is to define dynamic models that can be used to predict the behavior of modified strains and thus drive selection and genetic engineering.

4.3. Winemaking and improved strain selection

Yeasts and bacteria are essential for the winemaking process, and selection of strains based both on their efficiency and on the influence on the quality of wine is a subject of significant effort in the Aquitaine region. Unlike the species studied above, yeast and bacterial starters for winemaking cannot be genetically modified. In order to propose improved and more specialized starters, industrial producers use breeding and selection strategies.

Yeast starters from the *Saccharomyces* genus are used for primary, alcohol fermentation. Recent advances have made it possible to identify the genetic causes of the different technological differences between strains [55], [54], [53]. Manipulating the genetic causes rather than the industrial consequences is far more amenable to experimental development. An essential tool in identifying these genetic causes is comparative genomics.

Bacterial starters based on *Oenococcus oeni* are used in secondary, malolactic fermentation. Genetically, *O. oeni* presents a surprising level of intra-specific diversity, and clues that it may evolve more rapidly than expected. Studying the diversity of the emphO. oeni genomes has led to genetic tools that can be used to evaluate the predisposition of different strains to respond to oenological stresses. While identifying particular genes has been the leading strategy up to now, recently a new strategy based on comparative genomics has been undertaken to understand the impact and mechanisms of genetic diversity [20], [23], [21], [32], [36], [28] [3].

Starting from historical collaborations by Pascal Durrens and Elisabeth Bon with partners from the Institute for Wine and Vine Sciences in Bordeaux (ISVV), and local industry, and in the framework of an effective partnership, we apply our tools to large-scale comparative genomics of yeast and bacterial starters in winemaking.

4.4. Knowledge bases for molecular tools

Affinity binders are molecular tools for recognizing protein targets, that play a fundamental in proteomics and clinical diagnostics. Large catalogs of binders from competing technologies (antibodies, DNA/RNA aptamers, artificial scaffolds, etc.) and Europe has set itself the ambitious goal of establishing a comprehensive, characterized and standardized collection of specific binders directed against all individual human proteins, including variant forms and modifications. Despite the central importance of binders, they presently cover only a very small fraction of the proteome, and even though there are many antibodies against some targets (for example, >900 antibodies against p53), there are none against the vast majority of proteins. Moreover, widely accepted standards for binder characterization are virtually nonexistent.

Alongside the technical challenges in producing a comprehensive binder resource are significant logistical challenges, related to the variety of producers and the lack of reliable quality control mechanisms. As part of the ProteomeBinders and Affinomics projects, MAGNOME works to develop knowledge engineering techniques for storing, exploring, and exchanging experimental data used in affinity binder characterization. This work involves databases and tools for molecular interaction data [42] [45], standards for data exchange between peers [44], [49], [41] and reporting standards [4] [58].

<div align="center">

MNEMOSYNE Team

</div>

4. Application Domains

4.1. Overview

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

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

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

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

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

4. Application Domains

4.1. Resource Allocation and Scheduling

4.1.1. Project-team positioning

CEPAGE has undertaken tasks related to the *high level modeling* of heterogeneous networks, both at logical level (overlay networks design) and performance level (latency, bandwidth prediction, connectivity artifacts) in order to optimize tasks such as *resource allocation* and *scheduling* of computations and communications. Objectives include replica placement, broadcasting (streaming) of large messages, independent tasks scheduling and optimization of OLAP databases. Such problems have received at lot of attention in research centers in the USA (Armherst, Colorado, ...), in Spain (Madrid), Poland (Wroclaw), Germany (Dortmund), and others. Papers on algorithmic aspects of platform modeling, scheduling and resource allocation appear at parallel processing conferences and journals in Parallel and Distributed Computing (IPDPS, EuroPar, HIPC, SPAA, IEEE TPDS, JPDC) and members of CEPAGE are strongly involved in many of these events (IPDPS, EuroPar, TPDS) as well as helping to animate well-established specialized workshops, such as HCW and HeteroPar.

Within Inria, studies on overlay networks are performed in the ASAP and GANG projects, and studies related to scheduling and resource allocation are done within the ROMA and the MOAIS projects (and to some extent within ALGORILLE).

4.1.2. Scientific achievements

The approach followed in the CEPAGE project, and our main originality, is to consider the whole chain, from gathering actual data on the networks to platform modeling and complexity analysis. Indeed, many complexity analysis studies are performed on models whose parameters cannot actually be evaluated (this applies, for instance, to all algorithms that assume that the topology of a platform running over the Internet is known in advance) and many platform models are intractable from an algorithmic perspective (this applies, for instance, to all models that represent latencies or bandwidths between all pairs of nodes as a general matrix). Our general goal is to provide models whose parameters can be evaluated at runtime using actual direct measurements, to propose algorithms whose worst-case (or average-case) behavior can be proved for this model, and finally to evaluate the whole chain (model + algorithm + implementation).

From an applicative perspective, in the framework of the PhD Thesis of Hejer Rejeb, we have considered several storage and resource allocation problems in collaboration with Cyril Banino-Rokkones at Yahoo! Trondheim (dealing with actual datasets enabled us to improve known approximation results in this specific context). We have in particular studied the modeling of TCP mechanism for handling contentions and its influence on the performance of several scheduling algorithms and advocated the use of QoS mechanisms for prescribed bandwidth sharing (IPDPS 2010 [67], ICPADS 2008 [52], AlgoTel 2009 [64], ICPADS 2009 [63], PDP 2010 [65]). In the PhD thesis of Hubert Larchevêque, we have considered the problem of aggregating resources (or placing replicas) in a distributed network (Sirocco 2008 [54], Opodis 2008 [55], ICPP 2011 [60], AlgoTel 2011 [56]) so that each group satifies some properties (in terms of aggregated memory, CPU and maximal distance in terms of latency within a group). We proved several multi-criteria approximation results for this problem, and we compared several embedding tools (Vivaldi, Sequoia) in the context of resource aggregation. For these applications, we have also provided when possible distributed algorithms based on sophisticated overlay networks, in particular in order to deal with heterogeneity (IPDPS 2008 [61]). In the PhD Thesis of Przemyslaw Uznanski, we focus on the design of efficient streaming and broadcasting strategies, in particular in presence of connectivity artifacts like firewalls (IPDPS 2010 [62], ICPADS 2011 [59]). We have also worked on establishing under the bounded multiport model several new complexity results for classical distributed computing models such as divisible load theory (HCW 2008 [57], IPDPS 2008 [107], IPDPS 2012 [58]) that have been later extended to Continuous Integration (HCW 2012 [53]).

In the context of database query optimization, materializing some queries results for optimization is a standard solution when execution time performance is crucial. In the datacube context, the problem has been studied for a long time under the storage space limit constraint. Here also, we were able to reformulate this problem by considering instead the execution time as the hard constraint while the objective is to reduce the storage space. Even if the problem turns to be NP-hard, this reformulation allowed us to provide effective approximate solutions with both space and performance bounded guarantees (EDBT 2009 [96]). Moreover, reducing the storage space tends to reduce the maintenance time since the latter is linearly proportional to the former. Finally, we characterized the minimal number of updates to be performed before performance becomes no more guaranteed and a new solution must be recomputed (ADBIS 2008 [97]). One of the key concepts we used for solving this problem was that of a *border*. It turns out that this notion is equivalent to e.g., maximal frequent itemsets or minimal functional dependencies extensively studied by data mining community. In contrast to all previous proposals, we proposed the only parallel algorithm computing these borders with a speed-up guarantee regarding the number of processing units (CIKM 2011 [95]). Besides the analytical study, its implementation in maximal frequent itemset mining outperforms state of the art implementations (see Section 5.1).

To achieve these results, our efforts have also focused on analyzing and building realistic datasets (AlgoTel 2012 [86]) and proposing data analysis results for specific distributions (ISAAC 2011 [48]). On the modeling side, in general, for bandwidth and contention modeling, we have proved that the bounded multi-port model (where each node is associated to an incoming bandwidth, an outgoing bandwidth and a maximal number of simultaneous TCP connexions) is both implementable, realistic and tractable (EuroPar 2011 [66]). In particular, we have proved in strongly different contexts (allocation of virtual machines to physical machines, overlay design for broadcasting, server allocation for volunteer computing) that the use of resource augmentation enables to obtain quasi-optimal results. All our modeling efforts and algorithms have been included into the SimGRID Software (http://simgrid.gforge.inria.fr), which enables us both to compare several algorithms under the same exact conditions and to compare the results obtained with several communication models (see Section 5.1)..

Perspectives: We believe that our approach based on sound models, approximation algorithms for these models, followed by experimental validation is a strong one and we intend to continue in this direction in the following years. Our goal of designing realistic solutions pushes towards considering average case analysis of our algorithms, as well as robust optimization techniques. Furthermore, the recent strong interest in Cloud systems from the community entices us to use our expertise in resource allocation for the optimization of Cloud systems, both from the provider and from the user points of view. We already have some interesting contacts with local companies to share start collaborating on these topics. In this context, reliability issues are very important, and we believe that robust optimization is a very relevant approach for these problems.

4.2. Compact Routing

4.2.1. *Project-team positioning*

In this axis, CEPAGE mainly works on the design on distributed and light data structures. One of the techniques consists in summarizing the topology and metric of the networks allowing to route or to approximate the original distances within the network. Such structures, often called *spanners*, does not require the storage of all the original network links. Then we get economic distributed data structures that can be updated without a high communication cost. Our main collaborations are done with the best specialists world-wide, in particular: Israel (Weizmann), USA (MIT, Microsoft, Chicago), Belgium (Alcatel Lucent-Bell), France (Paris, Nice).

Algorithms and Routing are also intensively studied in research labs in the USA (CAIDA). Our contributions appear regularly at all of the major conferences in Distributed Computing (PODC, DISC, SPAA), as well as at top venues with a more general algorithmic audience (STOC, SODA, ICALP, ESA). Members of CEPAGE actively participate in these events (ICALP 2010 and DISC 2009 were organized by members of CEPAGE).

Within Inria, studies of mobile agents are also performed in the GANG project and to some extent also within MASCOTTE within the european project EULER.

4.2.2. Scientific achievements

There are several techniques to manage sub-linear size routing tables (in the number of nodes of the platform) while guaranteeing almost shortest paths. Some techniques provide routes of length at most $1 + \epsilon$ times the length of the shortest one while maintaining a poly-logarithmic number of entries per routing table. However, these techniques are not universal in the sense that they apply only on some class of underlying topologies. Universal schemes exist. Typically they achieve $O(\sqrt{n})$-entry local routing tables for a stretch factor of 3 in the worst case. Some experiments have shown that such methods, although universal, work very well in practice, in average, on realistic scale-free or existing topologies.

The space lower bound of $O(\sqrt{n})$-entry for routing with *multiplicative* stretch 3 is due to the existence of dense graphs with large girth. Dense graphs can be sparsified to subgraphs (spanners), with various stretch guarantees. There are spanners with *additive* stretch guarantees (some even have constant additive stretch) but only very few additive routing schemes are known.

In (SPAA 2012 [90]), we give reasons why routing in unweighted graphs with *additive* stretch is difficult in the form of space lower bounds for general graphs and for planar graphs. On the positive side, we give an almost tight upper bound: we present the first non-trivial compact routing scheme with $o(\lg^2 n)$-bit addresses, *additive* stretch $O(n^{1/3})$, and table size $O(n^{1/3})$ bits for planar graphs.

We have recently considered the *forbidden-set* extension of distance oracles and routing schemes. Given an arbitrary set of edge/node failure F, a source s and a target t such that $s, t \notin F$, the goal is to route (or evaluate the distance) between s and t in the graph $G \smallsetminus F$, so avoiding F. The classical problem is for $F = \varnothing$. This extension is considered as a first step toward fully dynamic data-structures, a challenging goal. For graphs of low doubling dimension we have shown in (PODC 2012 [47]) that it is possible to route from s to t in $G \smallsetminus F$ with stretch $1 + \epsilon$, for all s, t, F, given poly-logarithmic size labels of all the nodes invoked in the query (s, t, F). This has been generalized to all planar graphs achieving similar stretch and label size performances. As a byproduct we have designed a fully dynamic algorithm for maintaining $1 + \epsilon$ approximate distances in planar graphs supporting edge/node addition/removal within update and query time $\sqrt{n}$ in the worst-case (STOC 2012 [46]).

Θ_k-graphs are geometric graphs that appear in the context of graph navigation. The shortest-path metric of these graphs is known to approximate the Euclidean complete graph up to a factor depending on the cone number k and the dimension of the space. We have introduced in (WG 2010 [68]) a specific subgraph of the Θ_6-graph defined in the 2D Euclidean space, namely the half-Θ_6-graph, composed of the even-cone edges of the Θ_6-graph. Our main contribution is to show that these graphs are exactly the TD-Delaunay graphs, and are strongly connected to the geodesic embeddings of orthogonal surfaces of coplanar points in the 3D Euclidean space. We also studied the asymptotic behavior of these spanners (*Adv. in Appl. Proba.* [105]) and in collaboration with Ljubomir Perković, we worked on the question of bounded degree planar spanner. We proposed an algorithm that computes a plane 6-spanner of degree at most 6 in (ICALP 2010 [69]). The previous best bound on the maximum degree for constant stretch plane spanners was only 14.

In order to cope with network dynamism and failures, and motivated by multipath routing, we introduce a multi-connected variant of spanners. For that purpose we introduce in (OPODIS 2011 [91]) the p-multipath cost between two nodes u and v as the minimum weight of a collection of p internally vertex-disjoint paths between u and v. Given a weighted graph G, a subgraph H is a p-multipath s-spanner if for all u, v, the p-multipath cost between u and v in H is at most s times the p-multipath cost in G. The s factor is called the stretch. Building upon recent results on fault-tolerant spanners, we show how to build p-multipath spanners of constant stretch and of $O(n^{1+1/k})$ edges, for fixed parameters p and k, n being the number of nodes of the graph. Such spanners can be constructed by a distributed algorithm running in $O(k)$ rounds. Additionally, we give an improved construction for the case $p = k = 2$. Our spanner H has $O(n^{3/2})$ edges and the p-multipath cost in H between any two node is at most twice the corresponding one in G plus $O(W)$, W being the maximum edge weight.

We also worked on compact coding in data warehouses: in order to get quick answer in large data, we have to estimate, select and materialize (store) partial data structures. We got several solutions with a prescribed

guarantee in different models for the following problems: view size estimation with small samples, view selection, parallel computation of frequent itemsets. In (*Theor. Comp. Sci.* [94]) a new algorithm that allow the administrator or user of a DBMS to choose which part of the data cube to optimize (known as the *the views selection problem*), that takes as input a fact table and computes a set of views to store in order to speed up queries.

Perspectives: The compact coding activity in data-warehouse is promising since the amount of data collected keeps on increasing and being able to answer in real-time complex requests (data mining) is still challenging.

Some robust data structures already exist which, given a small number of k changes of topology or k faults, tolerate these faults, i.e., alternative routes with bounded stretch can be provided without any updates. This is a first step toward dynamic networks but the updates of these data structures are currently still quite complicated with a high communication cost.

4.3. Mobile Agents

4.3.1. *Project-team positioning*

CEPAGE has undertaken tasks related to the design of algorithms which control the behavior of so called *mobile agents*, moving around a network or a geometric environment, with the goal of achieving a specified objective. Objectives of central importance to the study include: exploration of unknown environments, terrain patrolling, network maintenance, and coordination of activities with other agents. Such problems have in recent years been the object of interest of numerous research teams working on Distributed Computing worldwide, in particular, at research centers in Canada (Quebec), Israel (Tel Aviv, Haifa), France (Paris, Marseille), the UK (London, Liverpool), and Switzerland (Zurich). Algorithms for mobile agents in social networking applications are also intensively studied in research labs in the USA (Stanford, Facebook). Papers on mobile agents appear regularly at all of the major conferences in Distributed Computing (PODC, DISC, SPAA), as well as at top venues with a more general algorithmic audience (SODA, ICALP, ESA). Members of CEPAGE actively participate in these events, and are also a recognizable part of the European community focused around mobile agents, helping to animate well-established specialized conferences, such as SIROCCO and OPODIS.

Within Inria, studies of mobile agents are also performed in the GANG project, and to some extent also within MASCOTTE. CEPAGE has active research links with both of these teams.

4.3.2. *Scientific achievements*

The work of CEPAGE has focused on contributing new decentralized algorithms for controlling mobile entities known as *agents*, deployed in unknown environments. We mainly considered the network setting, in which agents moving around the nodes of the network graph may be used to analyze the structure of the network and to perform maintenance tasks, such as detecting dynamic faults, improving/monitoring dissemination of information, etc. Our theoretical studies focused on designing new strategies for controlling the behavior of agents and answering crucial questions concerning the feasibility of solving fundamental problems, subject to different model assumptions and constraints on the knowledge and computational power of agents.

One major line of our research focused on the so called *anonymous graph model* in which an agent is unable to determine the identifier of the node of its current location, but can only see a local ordering of the links around it. Such a study is motivated e.g. by scenarios in which the identifiers of nodes may be too large for the agent to process using its bounded resources, or may change in time. In this model, we studied two of the most fundamental problems: that of traversing all of the nodes of the network (exploration) and of meeting another agent in the network (rendezvous), so as to coordinate with it. Our contributions include a precise characterization of the space requirements for agents solving both of these problems deterministically: exploration in (*Trans. Alg.* 2008 [73]) and rendezvous in (*Dist. Comp.* 2012 [81]), in a paper presented at the Best Paper Session of PODC 2010. We have also studied fast solutions for specific scenarios of the rendezvous problem (DISC 2010 [49], DISC 2011 [74], SPAA 2012 [82]) and the problem of approximate map construction within an anonymous graph (OPODIS 2010 [71]). A separate problem, intensively studied in recent years by several research teams, concerns the exploration of a network with pre-configured ports

so as to assist the agent. In our work on the topic, our team has proposed several new techniques for graph decomposition, leading in particular to the shortest currently known strategies of periodic exploration for both the case of memoryless (*Algorithmica* 2012 [101]) and small-memory agents (SIROCCO 2009 [77]).

A closely related line of research was devoted to the design of network exploration strategies which guarantee a fast and fair traversal of all the nodes, making use of agents with extremely restricted capabilities. Such strategies were inspired by the random walk, but had the additional advantage of deterministic and desirable behavior in worst-case scenarios. We presented a series of results in the area at notable conferences, involving both the design of new exploration strategies (ICALP 2009 [75]) and completely new insights into previously known approaches such as the so called "rotor-router model" (DISC 2009 [50], OPODIS 2009 [51]). All of the proposed algorithms were shown to be viable alternatives to the random walk, competing in terms of such parameters as cover time, steady-state exploration frequency, and stabilization in the event of faults.

Our efforts have also focused on the theory of coordinating activities of large groups of agents. We have conducted pioneering work in the so called look-compute-move model in networks, in which extremely restricted (asynchronous and oblivious) agents, relying on snapshot views of the system, are nevertheless able to perform useful computational tasks. Our solutions to the problems of collective exploration in trees (*Theor. Comp. Sci.* 2010 [88]) and gathering agents on a ring (*Theor. Comp. Sci.* 2008 [99] and 2010 [98]) have sparked a long line of follow-up research, accumulating more than 120 citations in total (according to Google Scholar). In a slightly different scenario, we have considered computations with teams of agents whose task is to collaboratively detect and mark potentially dangerous (faulty) links of the network, called "black holes", which are capable of destroying agents which enter them. We have provided important contributions to the theory of black hole search in both undirected (SIROCCO 2008 [76], DISC 2008 [89]) and directed (*Theor. Comp. Sci.* [102]) graphs.

It is expected that the mobile agent theme of CEPAGE will give rise to 2 PhD theses. In 2013, Ahmed Wade will defend his thesis on mobile agent protocols for dynamic networks, whereas in 2014 Dominik Pajak will defend his thesis on multi-agent protocols for efficient graph exploration. Our scientific interests also include mobile agent protocols for geometric applications, more remote from the central themes of CEPAGE, but having extensive applications in robotics (providing protocols, e.g., for efficient patrolling and guarding of terrains, traversing terrains using groups of robots, etc.). We have already published several papers in this area (SIROCCO 2010 [79], SWAT 2010 [80], ESA 2011 [78]), building up the theoretical fundamentals of a new field, and already attracting the attention of a wider community of researchers working in robotics and AI.

Perspectives: Our goal is to explore applications of mobile agent techniques in domains of growing importance, namely, social networks and robotics. We are currently discussing applications of our techniques in problems of brand recognition on the web with a local industrial partner (Systonic KeepAlert), and other companies (through our research collaborators in Liverpool). We intend to undertake collaboration with European/American research labs and industrial partners.

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

4. Application Domains

4.1. Introduction

Currently, we have one major application which is material physics, and for which we contribute to all steps that go from modelling aspects to the design and the implementation of very efficient algorithms and codes for very large multi-scale simulations. Moreover, we apply our algorithmic research about linear algebra (see Section 3) in the context of several collaborations with industrial and academic partners. Our high performance libraries are or will be integrated in several complex codes and will be used and validated for very large simulations.

4.2. Material physics

Participants: Bérenger Bramas, Olivier Coulaud, Aurélien Esnard, Pierre Fortin, Luc Giraud, Jean Roman.

Due to the increase of available computer power, new applications in nano science and physics appear such as study of properties of new materials (photovoltaic materials, bio- and environmental sensors, ...), failure in materials, nano-indentation. Chemists, physicists now commonly perform simulations in these fields. These computations simulate systems up to billion of atoms in materials, for large time scales up to several nanoseconds. The larger the simulation, the smaller the computational cost of the potential driving the phenomena, resulting in low precision results. So, if we need to increase the precision, there is two ways to decrease the computational cost. In the first approach, we improve classical methods and algorithms and in the second way, we will consider a multiscale approach.

Many applications in material physics need to couple several models like quantum mechanic and molecular mechanic models, or molecular and mesoscopic or continuum models. These couplings allow scientists to treat larger solids or molecules in their environment. Many of macroscopic phenomena in science depend on phenomena at smaller scales. Full simulations at the finest level are not computationally feasible in the whole material. Most of the time, the finest level is only necessary where the phenomenon of interest occurs; for example in a crack propagation simulation, far from the tip, we have a macroscopic behavior of the material and then we can use a coarser model. The idea is to limit the more expensive level simulation to a subset of the domain and to combine it with a macroscopic level. This implies that atomistic simulations must be speeded up by several orders of magnitude.

We will focus on two applications; the first one concerns the computation of optical spectra of molecules or solids in their environment. In the second application, we will develop faster algorithms to obtain a better understanding of the metal plasticity, phenomenon governing by dislocation behavior. Moreover, we will focus on the improvement of the algorithms and the methods to build faster and more accurate simulations on modern massively parallel architectures.

4.2.1. *Hybrid materials*

There is current interest in hybrid pigments for cosmetics, phototherapy and paints. Hybrid materials, combining the properties of an inorganic host and the tailorable properties of organic guests, particularly dyes, are also of wide interest for environmental detection (oxygen sensors) and remediation (trapping and elimination of dyes in effluents, photosensitised production of reactive oxygen species for reduction of air and water borne contaminants). A thorough understanding of the factors determining the photo and chemical stability of hybrid pigments is thus mandated by health, environmental concerns and economic viability.

Many applications of hybrid materials in the field of optics exploit combinations of properties such as transparency, adhesion, barrier effect, corrosion, protection, easy tuning of the colour and refractive index, adjustable mechanical properties and decorative properties. It is remarkable that ancient pigments, such as Maya Blue and lacquers, fulfill a number of these properties. This is a key to the attractiveness of such materials. These materials are not simply physical mixtures, but should be thought of as either miscible organic and inorganic components, or as a heterogeneous system where at least one of the component exhibits a hierarchical order at the nanometer scale. The properties of such materials no longer derive from the sum of the individual contributions of both phases, since the organic/inorganic interface plays a major role. Either organic and inorganic components are embedded and only weak bonds (hydrogen, van der Waals, ionic bonds) give the structure its cohesion (class I) or covalent and iono-covalent bonds govern the stability of the whole (class II).

These simulations are complex and costly and may involve several length scales, quantum effects, components of different kinds (mineral-organic, hydro-philic and -phobic parts). Computer simulation already contributes widely to the design of these materials, but current simulation packages do not provide several crucial functions, which would greatly enhance the scope and power of computer simulation in this field.

The computation of optical spectra of molecules and solids is the greatest use of the Time Dependent Density Functional Theory (TDDFT). We compute the ground state of the given system as the solution of the Kohn-Sham equations (DFT). Then, we compute the excited states of the quantum system under an external perturbation - electrical field of the environment - or thanks to the linear theory, we compute only the response function of the system. In fact, physicists are not only interesting by the spectra for one conformation of the molecule, but by an average on its available configurations. To do that, they sample the trajectory of the system and then compute several hundred of optical spectra in one simulation. But, due to the size of interesting systems (several thousands of atoms) and even if we consider linear methods to solve the Kohn-Sham equations arising from the Density Functional Theory, we cannot compute all the system at this scale. In fact, such simulations are performed by coupling Quantum mechanics (QM) and Molecular mechanic (MM). A lot of works are done on the way to couple these two scales, but a lot of work remains in order to build efficient methods and efficient parallel couplings.

The most consuming time in such coupling is to compute optical spectra is the TDDFT. Unfortunately, examining optical excitations based on contemporary quantum mechanical methods can be especially challenging because accurate methods for structural energies, such as DFT, are often not well suited for excited state properties. This requires new methods designed for predicting excited states and new algorithms for implementing them. Several tracks will be investigated in the project:

- Typically physicists or chemists consider spectral functions to build a basis (orbital functions) and all the computations are performed in a spectral way. Due to our background, we want to develop new methods to solve the system in the real space by finite differences or by wavelets methods. The main expectation is to construct error estimates based on for instance the grid-size h parameter.

- For a given frequency in the optical spectra, we have to solve a symmetric non Hermitian system. With our knowledge on linear solvers, we think that we can improve the methods commonly used (Lanczos like) to solve the system (see Section 3.3).

- Improving the parallel coupling is crucial for large systems because the computational cost of the atomic and quantum models are really different. In parallel we have the following order of magnitude: one second or less per time step for the molecular dynamics, several minutes or more for the DFT and the TDDFT. The challenge to find the best distribution in order to have the same CPU time per time step is really important to reach high performance. Another aspect in the coupling is the coupling with the visualization to obtain online visualization or steerable simulations. Such steerable simulations help the physicists to construct the system during the simulation process by moving one or a set of molecules. This kind of interaction is very challenging in terms of algorithmic and this is a good field for our software platform EPSN .

4.2.2. Material failures

Another domain of interest is the material aging for the nuclear industry. The materials are exposed to complex conditions due to the combination of thermo-mechanical loading, the effects of irradiation and the harsh operating environment. This operating regime makes experimentation extremely difficult and we must rely on multi-physics and multi-scale modelling for our understanding of how these materials behave in service. This fundamental understanding helps not only to ensure the longevity of existing nuclear reactors, but also to guide the development of new materials for 4th generation reactor programs and dedicated fusion reactors. For the study of crystalline materials, an important tool is dislocation dynamics (DD) modelling. This multiscale simulation method predicts the plastic response of a material from the underlying physics of dislocation motion. DD serves as a crucial link between the scale of molecular dynamics and macroscopic methods based on finite elements; it can be used to accurately describe the interactions of a small handful of dislocations, or equally well to investigate the global behavior of a massive collection of interacting defects.

To explore, i.e., to simulate these new areas, we need to develop and/or to improve significantly models, schemes and solvers used in the classical codes. In the project, we want to accelerate algorithms arising in those fields. We will focus on the following topics (in particular in the starting OPTIDIS ANR-COSINUS project in collaboration with CEA Saclay, CEA Ile-de-france and SIMaP Laboratory in Grenoble) in connection with research described at Sections 3.4 and 3.5 .

- The interaction between dislocations is long ranged ($O(1/r)$) and anisotropic, leading to severe computational challenges for large-scale simulations. In dislocation codes, the computation of interaction forces between dislocations is still the most CPU time consuming and has to be improved to obtain faster and more accurate simulations.

- In such simulations, the number of dislocations grows while the phenomenon occurs and these dislocations are not uniformly distributed in the domain. This means that strategies to dynamically construct a good load balancing are crucial to acheive high performance.

- From a physical and a simulation point of view, it will be interesting to couple a molecular dynamics model (atomistic model) with a dislocation one (mesoscale model). In such three-dimensional coupling, the main difficulties are firstly to find and characterize a dislocation in the atomistic region, secondly to understand how we can transmit with consistency the information between the two micro and meso scales.

4.3. Application framework customers of high performance linear algebra solvers

Participants: Emmanuel Agullo, Luc Giraud, Abdou Guermouche, Stojce Nakov, Jean Roman, Xavier Vasseur.

We are currenlty collaborating with various research groups involved in geophysics, electromagnetics and structural mechanics. For all these application areas, the current bottleneck is the solution of huge sparse linear systems often involving multiple right-hand sides either available simultaneously or given in sequence. The robustness, efficiency and scalability of the numerical tools designed in Section 3.3 will be preliminary investigated in the parallel simulation codes of these partners.

For the solution of large systems arsing from PDE discretization, the geometric full multigrid technique based on a few levels in the grid hierarchy and an efficient parallel sparse direct solver on the coarsest level can be considered. Originally developed for 3D Maxwell solution in collaboration with CEA-CESTA, the approach can be extended to other application fields.

Many simulation codes need the solution with simultaneous right-hand sides but also with right-hand sides given in sequence. The first situation arises in RCS calculations, but is generic in many parametric studies, while the second one comes from the nature of the solver such as implicit time tepping schemes or inverse iterations. Many of the numerical approaches and possible outcoming software are well suited to tackle these challenging problems.

On more academic sides, some ongoing collaborations with other Inria EPIs will be continued and others will be started. In collaboration with the NACHOS Inria project team, we will continue to investigate the use of efficient linear solvers for the solution of the Maxwell equations in the time and frequency domains where discontinuous Galerkin discretizations are considered. Additional funding will be sought out in order to foster this research activity in connection with actions described in Section 3.3 .

The efficient solution of linear systems strongly relies on the activities described in Section 3.2 (e.g. complex load balancing problem) and in Section 3.3 (for the various parallel linear algebra kernels).

4.4. Scalable numerical schemes for scientific applications

Participants: Olivier Coulaud, Yohann Dudouit, Luc Giraud, Guillaume Latu, Alexis Praga, Jean Roman, Pablo Salas Medina, Xavier Vasseur.

We are also collaborating with application research group to design or improve numerical schemes in the view of large scale parallel simulations.

Seismic wave propagation in heterogeneous media requires to properly capture the local heterogeneity and consequently requires locally refined meshes. In close collaboration with TOTAL we study new parallelizable schemes for the solution of the elastodynamic system with local spatial refinments based on discontinuous Galerkin techniques. The objective is to design novel parallel scalable implementations for large 3D simulations. A second work is currently carried on with TOTAL for Seismic modeling and Reverse Time Migration (RTM) based on the full wave equation discretization. These tools are of major importance since they give an accurate representation of complex wave propagation areas. Unfortunately, they are highly compute intensive. To address this challenge we have designed a fast parallel simulator that solves the acoustic wave equation on a `GPU` cluster.

Thermoacoustic instabilities are an important concern in the design of gas turbine combustion chambers. Most modern combustion chambers have annular shapes and this leads to the appearance of azimuthal acoustic modes. These modes are often powerful and can lead to structural vibrations being sometimes damaging. Therefore, they must be identified at the design stage in order to be able to eliminate them. However, due to the complexity of industrial combustion chambers with a large number of burners, numerical studies of real configurations are a challenging task. Such a challenging calculations performed in close collaboration with the Computational Fluid Dynamic project at CERFACS.

The chemistry and transport models (CTM) play a central role in global geophyscal models. The solution of the CTM represents up-to 50 % on the computing ressources involved in global geophyscal simulations. Therefore, the availability of efficient scalable parallel numerical schemes on emerging and future supercomputers is crucial. The purpose of this research activity is to study, design and implement novel numerical schemes following the work initiated by D. Cariolle in the framework of the ANR Solstice project. Alexi Praga, PhD hired by CERFACS, is conducing this research action under the joint supervision of `HiePACS` and the Aviation and Environment project at CERFACS in close collaboration with CNRM/Meteo-France.

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

4. Application Domains

4.1. Introduction

Building on our previous work, we are studying software development in the context of communication services, in their most general forms. That is, going beyond human-to-human interactions, and covering human-to-machine and machine-to-machine interactions. Software systems revolving around such forms of communications can be found in a number of areas, including telephony, pervasive computing, and assisted living; we view these software systems as coordinating the communication between networked entities, regardless of their nature, human, hardware or software. In this context, our three main application domains are pervasive computing, avionics and assisted living.

4.2. Pervasive Computing

Pervasive computing systems are being deployed in a rapidly increasing number of areas, including building automation and supply chain management. Regardless of their target area, pervasive computing systems have a typical architectural pattern. They aggregate data from a variety of distributed sources, whether sensing devices or software components, analyze a context to make decisions, and carry out decisions by invoking a range of actuators. Because pervasive computing systems are standing at the crossroads of several domains (*e.g.,* distributed systems, multimedia, and embedded systems), they raise a number of challenges in software development:

- *Heterogeneity.* Pervasive computing systems are made of off-the-shelf entities, that is, hardware and software building blocks. These entities run on specific platforms, feature various interaction models, and provide non-standard interfaces. This heterogeneity tends to percolate in the application code, preventing its portability and reusability, and cluttering it with low-level details.

- *Lack of structuring.* Pervasive computing systems coordinate numerous, interrelated components. A lack of global structuring makes the development and evolution of such systems error-prone: component interactions may be invalid or missing.

- *Combination of technologies.* Pervasive computing systems involve a variety of technological issues, including device intricacies, complex APIs of distributed systems technologies and middleware-specific features. Coping with this range of issues results in code bloated with special cases to glue technologies together.

- *Dynamicity.* In a pervasive computing system, devices may either become available as they get deployed, or unavailable due to malfunction or network failure. Dealing with these issues explicitly in the implementation can quickly make the code cumbersome.

- *Testing.* Pervasive computing systems are complicated to test. Doing so requires equipments to be acquired, tested, configured and deployed. Furthermore, some scenarios cannot be tested because of the nature of the situations involved (*e.g.,* fire and smoke). As a result, the programmer must resort to writing specific code to achieve ad hoc testing.

4.3. Assisted Living

Cognitive impairments (memory, attention, time and space orientation, *etc*) affect a large part of the population, including elderly, patients with brain injuries (traumatic brain injury, stroke, *etc*), and people suffering from cognitive disabilities, such as Down syndrome.

The emerging industry of digital assistive technologies provide hardware devices dedicated to specific tasks, such as a telephone set with a keyboard picturing relatives (http://www.doro.fr), or a device for audio and video communication over the web (http://www.technosens.fr). These assistive technologies apply a traditional approach to personal assistance by providing an equipment dedicated to a single task (or a limited set of tasks), without leveraging surrounding devices. This traditional approach has fundamental limitations that must be overcome to significantly improve assistive technologies:

- they are *not adaptable to one's needs*. They are generally dedicated to a task and have very limited functionalities: no networking, limited computing capabilities, a limited screen and rudimentary interaction modalities. This lack of functionality may cause a proliferation of devices, complicating the end-user life. Moreover, they are rarely designed to adapt to the cognitive changes of the user. When the requirements evolve, the person must acquire a new device.

- they are often *proprietary*, limiting innovation. As a result, they cannot cope with the evolution of users' needs.

- they have limited or *no interoperability*. As a result, they cannot rely on other devices and software services to offer richer applications.

To break this model, we propose to offer an assistive solution that is open-ended in terms of applications and entities. (1) An on-line catalog of available applications enables every user and caregiver to define personalized assistance in the form of an evolving and adapted set of applications; this catalog provides a community of developers with a mechanism to publish applications for specific daily-activity needs. (2) New types of entities can be added to a platform description to enhance its functionalities and extend the scope of future applications.

4.4. Avionics

In avionics, an aircraft can be seen as an environment full of sensors (*e.g.,* accelerometers, gyroscopes, and GPS sensors) and actuators (*e.g.,* ailerons and elevator trim). For example, a flight guidance system controls the aircraft using data produced by sensors. In a critical platform such as an aircraft, software systems have to be certified. Moreover the safety-critical nature of the avionics domain takes the form of stringent non-functional requirements, resulting in a number of challenges in software development:

- *Traceability.* Traceability is the ability to trace all the requirements throughout the development process. In the avionics certification processes, traceability is mandatory for both functional and non-functional requirements.

- *Coherence.* Functional and non-functional aspects of an application are inherently coupled. For example, dependability mechanisms can potentially deteriorate the overall performance of the application. The coherence of the requirements is particularly critical when the software evolves: even minor modifications to one aspect may tremendously impact the others, leading to unpredicted failures.

- *Separation of concerns.* Avionics platforms involve the collaboration of several experts (from low-level system to software, safety, QoS), making requirements traceability significantly more challenging. Providing development methodologies that allow a clear separation of concerns can tremendously improve traceability.

Our approach consists of enriching a design language with non-functional decalarations. Such declarations allow the safety expert to specify at design time how errors are handled, guiding and facilitating the implementation of error handling code. The design is also enriched with Quality of Service (QoS) declarations such as time constraints. For each of these non-functional declarations, specific development support can be generated. We have validated this approach by developing flight guidance applications for avionics and drone systems.

RUNTIME Project-Team

4. Application Domains

4.1. Application Domains

HPC, simulation

The RUNTIME group is working on the design of efficient runtime systems for parallel architectures. We are currently focusing our efforts on High Performance Computing applications that merely implement numerical simulations in the field of Seismology, Weather Forecasting, Energy, Mechanics or Molecular Dynamics. These time-consuming applications need so much computing power that they need to run over parallel machines composed of several thousands of processors.

Because the lifetime of HPC applications often spreads over several years and because they are developed by many people, they have strong portability constraints. Thus, these applications are mostly developed on top of standard APIs (e.g. MPI for communications over distributed machines, OpenMP for shared-memory programming). That explains why we have long standing collaborations with research groups developing parallel language compilers, parallel programming environments, numerical libraries or communication software. Actually, all these "clients" are our primary target.

Although we are currently mainly working on HPC applications, many other fields may benefit from the techniques developed by our group. Since a large part of our efforts is devoted to exploiting multicore machines and GPU accelerators, many desktop applications could be parallelized using our runtime systems (e.g. 3D rendering, etc.).

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

4. Application Domains

4.1. Application Domains

- **Personal robotics.** Many indicators show that the arrival of personal robots in homes and everyday life will be a major fact of the 21st century. These robots will range from purely entertainment or educative applications to social companions that many argue will be of crucial help in our aging society. For example, UNECE evaluates that the industry of entertainment, personal and service robotics will grow from $5.4Bn to $17.1Bn over 2008-2010. Yet, to realize this vision, important obstacles need to be overcome: these robots will have to evolve in unpredictable homes and learn new skills while interacting with non-engineer humans after they left factories, which is out of reach of current technology. In this context, the refoundation of intelligent systems that developmental robotics is exploring opens potentially novel horizons to solve these problems.

- **Human-Robot Collaboration.** Robots play a vital role for industry and ensure the efficient and competitive production of a wide range of goods. They replace humans in many tasks which otherwise would be too difficult, too dangerous, or too expensive to perform. However, the new needs and desires of the society call for manufacturing system centered around personalized products and small series productions. Human-robot collaboration could widen the use of robot in this new situations if robots become cheaper, easier to program and safe to interact with. The most relevant systems for such applications would follow an expert worker and works with (some) autonomy, but being always under supervision of the human and acts based on its task models.

- **Video games.** In conjunction with entertainment robotics, a new kind of video games are developing in which the player must either take care of a digital creature (e.g. Neopets), or tame it (e.g. Nintendogs), or raise/accompany them (e.g. Sims). The challenges entailed by programming these creatures share many features with programming personal/entertainment robots. Hence, the video game industry is also a natural field of application for FLOWERS.

- **Environment perception in intelligent vehicles.** When working in simulated traffic environments, elements of FLOWERS research can be applied to the autonomous acquisition of increasingly abstract representations of both traffic objects and traffic scenes. In particular, the object classes of vehicles and pedestrians are if interest when considering detection tasks in safety systems, as well as scene categories ("scene context") that have a strong impact on the occurrence of these object classes. As already indicated by several investigations in the field, results from present-day simulation technology can be transferred to the real world with little impact on performance. Therefore, applications of FLOWERS research that is suitably verified by real-world benchmarks has direct applicability in safety-system products for intelligent vehicles.

- **Automated Tutoring Systems.** Optimal teaching and efficient teaching/learning environments can be applied to aid teaching in schools aiming both at increase the achievement levels and the reduce time needed. From a practical perspective, improved models could be saving millions of hours of students' time (and effort) in learning. These models should also predict the achievement levels of students in order to influence teaching practices.

MANAO Team (section vide)

4. Application Domains

4.1. Application domains

Since our project aims at providing 3D digital worlds to all, including the general public, in order to stimulate understanding, learning, communication and creation, our scope of applications will naturally be the following one:

- **Culture and education:** We are convinced that a 3D digital world is a powerful media that may contribute to enhance understanding processes. For example, a museum would benefit from new 3D user interfaces allowing visitors to better understand complex content. Similarly, at school, this media has an extraordinary potential for enhancing learning. For example, a child being able to navigate in archaeological sites, or being able to manipulate by himself 3D molecules as described previously, will probably understand and learn things while having pleasure in interacting with the content.

- **Art:** We believe that 3D digital worlds may stimulate creativity, too. Our first investigations with music and drawings have shown that this media opens new possibilities for creation. The challenge here will be to design good interfaces that will allow artists to explore new dimensions. The user may be an experienced artist, or a three years old child who would express his creativity thought tools that go beyond papers and pens.

- **Healthcare:** People suffering from motor or cognitive impairments are one of the target populations of the Potioc project. Indeed, we believe that new interfaces that exploit 3D digital worlds may help people to overcome their disabilities. For example, someone with very reduced motor capabilities could benefit from BCI to explore a virtual museum, or a children having difficulties for concentrating may benefit from new 3D interactive systems.

- **Entertainment:** The objective of Potioc is to open 3D digital worlds to everyone by designing innovative interfaces driven by enjoyment. Consequently, the entertainment industry will be an obvious application domain where Potioc can contribute. This can be in the scope of videogames, entertainment parks, Web and TV of the future, applications for mobile devices, and so on.

Naturally, we will not necessarily address all these applications, and certainly not all at the same time. These areas just define the applicative scope of our work. As an example, much of our current work is targeted at artistic and entertainment applications, with VR-based musical performances, augmented paper-based drawing or BCI-based video games. We are also currently starting to conduct research on digital cities, in order to provide ordinary citizens suitable tools and UI to explore 3D content related to their city, such as 3D maps, information about population density or sound nuisance, among other. It should also be noted that our work might find applicative connexions outside these main application domains and benefit to a large range of academic and industrial areas, with which we could build relationships. For example, in the scope of medicine, new and easy to adopt user interfaces designed in Potioc could prove valuable for medical professionals as well, to better access and interact with biological 3D content (e.g., X-rays or MRI scans).