| 1. AMIB Project-Team                  | 4 |
| 2. AVIZ Project-Team                 | 9 |
| 3. BYMOORE Exploratory Action        | 12 |
| 4. COMETE Project-Team               | 14 |
| 5. COMMANDS Project-Team             | 17 |
| 6. DAHU Project-Team                 | 21 |
| 7. DEFI Project-Team                 | 23 |
| 8. DIGIPLANTE Team                   | 30 |
| 9. DISCO Team                        | 35 |
| 10. GALEN Team                       | 42 |
| 11. GECO Team                        | 45 |
| 12. GEOMETRICA Project-Team          | 48 |
| 13. GRAND-LARGE Project-Team         | 57 |
| 14. HIPERCOM Project-Team            | 67 |
| 15. IN-SITU Project-Team             | 77 |
| 16. LEO Team                         | 81 |
| 17. MAXPLUS Project-Team             | 89 |
| 18. MEXICO Project-Team              | 106 |
| 19. PARIETAL Project-Team            | 109 |
| 20. PARSIFAL Project-Team            | 118 |
| 21. PROVAL Project-Team              | 121 |
| 22. REGULARITY Team                  | 124 |
| 23. SECSI Project-Team               | 135 |
| 24. SELECT Project-Team              | 140 |
| 25. TANC Project-Team                | 143 |
| 26. TAO Project-Team                 | 145 |
| 27. TYPICAL Project-Team             | 152 |
5. New Results

5.1. RNA structures

5.1.1. RNA secondary structures: folding, design and evolution

In a collaboration with J. Waldispuhl (McGill, Canada) (Presented at the RECOMB’11 conference [32]), we used weighted grammatical models, introduced by members of the group [2], to perform an efficient exploration of the mutational landscape of RNA. We proposed an adaptive sampling algorithm, where weights were used to compensate an identified bias toward regions of higher GC-content within sampled sequences, thereby allowing for the exploration of more relevant portions of the evolutionary landscape. These adaptive sampling principles can be adapted into a method for the RNA design following similar principles. This constitutes a competitive alternative to local search strategies used by all existing tools for this problem. This work is ongoing as a collaboration with B. Berger group (MIT) and J. Waldispuhl (McGill, Canada), and a manuscript was recently submitted.

5.1.2. RNA knowledge-based potentials and 3D studies

We used the curated database of biologically interesting structures we have set up to perform a statistical analysis and developed knowledge-based potentials. The database server is available at http://csb.stanford.edu/rna.

We obtained RNA knowledge-based potentials that now performs well at different representation levels. They can be used in three well-known Molecular Dynamics (MD) and modeling software suites ENCAD [44], GROMACS (v3 and 4) [45] and MOSAICS [41] and are available for the community. The study we performed on a large number of new decoys showed that our potential outperforms Rosetta RNA scoring function [37] which is the gold standard. We show that not having correction terms for base-stacking and pairing can be of advantage when modelling loops at high resolution. The study was welcomed by the RNA community and published in RNA Journal (IF 6.5) [8].

We also refined the mixture model strategy we developed for building knowledge-based potentials. In collaboration with O. Schwander at LIX, we compared different mixture models: Dirichlet Process Mixture models (DPM), Kernel Density Estimation (KDE) models, Expectation Maximization models (MM) with different number of components (including a simplified version based on a post-processing step using K-Means). We showed that the Dirichlet Process Mixtures (DPM) is a good tradeoff despite its longer precomputation time as it provides a smooth potential having relatively few components. This study was presented at the MCMMB’11 conference and was submitted as a journal paper.

This work was done in collaboration with A. Sim, X. Huang an M. Levitt (Stanford University - GNAPI Associate team).

5.2. Proteins structures

5.2.1. Protein sequence alignment

In comparative protein modeling, the quality of a template model depends heavily on the quality of the initial alignment between a given protein with unknown structure to various template proteins, whose tertiary structure is available in the Protein Data Bank (PDB). Although pairwise sequence alignment has been solved for more than three decades, there remains a large discrepancy between the accuracy of the best sequence alignment between two amino acid sequences, as produced by the Needleman-Wunsch or Smith-Waterman algorithms, and that of the best structural alignment between two protein X-ray structures, as produced by the software DALI, CE, TOPOFIT, etc. To improve the quality of initial alignments in template modeling, one can integrate valuable information from an ensemble of generated suboptimal alignments, that is alignments whose score is below the best possible score. In a collaboration with P. Clote (Boston College/DIGITEO) [26], we presented a novel algorithm to produce suboptimal pairwise alignments.
5.2.2. Protein-protein interaction:

A protein-protein docking procedure traditionally consists in two successive tasks: a search algorithm generates a large number of candidate solutions, and then a scoring function is used to rank them in order to extract a native-like conformation. We have already demonstrated that using Voronoi constructions and a defined set of parameters, we could optimize an accurate scoring function. However, the precision of such a function is still not sufficient for large-scale exploration of the interactome.

Another geometric construction was also tested: the Laguerre tessellation. It also allows fast computation without losing the intrinsic properties of the biological objects. Related to the Voronoi construction, it was expected to better represent the physico-chemical properties of the partners. In , we present the comparison between both constructions.

We also worked on introducing a hierarchical analysis of the original complex three-dimensional structures used for learning, obtained by clustering. Using this clustering model we can optimize the scoring functions and get more accurate solutions. This scoring function has been tested on CAPRI scoring ensembles, and an at least acceptable conformation is found in the top 10 ranked solutions in all cases. This work was part of the thesis of Thomas Bourquard, defended in 2009.

A strong emphasis was recently made on the design of efficient complex filters. To achieve this goal, we focused on the use of collaborative filtering methods state of the art machine learning approaches combined with our genetic algorithm [ 9 ].

We have also proposed an approach that improves the predictions made by HEX, a state-of-the art docking tool developed by INRIA Nancy. We applied Voronoi fingerprint to the output of HEX and learn how to rank them, and we have tested new ranking strategies. The obtained ranking improve the initial ranking of HEX [ 33 ], [ 23 ].

We also decided to extend these techniques to the analysis of protein-nucleic acid complexes. The first preliminary developments and tests were performed by Adrien Guilhot during his M1 internship for two months.

5.2.3. Transmembrane β-barrels:

We have recently proposed an algorithm [ 31 ] that classifies Transmembrane β-Barrel Proteins (TMB) and predicts their structure. It first uses a simple probabilistic model to filter out the proteins and strands which are not beta-barrel. Then, we build a graph-theoretic model to fold into the super-secondary structure via dynamic programming. This step runs in \( O(n^3) \) time for the common up-down topology, and at most \( O(n^5) \) for the Greek key motifs, where \( n \) is the number of amino acids. Finally a predicted three-dimensional structure is built from the geometric criteria. If the pseudoenergy is insufficient, the protein is classified as a non-TMB protein.

We have tested this approach on TMB and non-TMB proteins for classification and structure prediction. We tested classification on a dataset of 14238 proteins including 48 TMB and 14190 non-TMB proteins. Our classification results are very accurate and comparable to other algorithms [ 21 ], [ 5 ]. Especially, our PPV, MCC and F-Scores are second only to a very recent algorithm by Freeman and Wimley [ 39 ], which relies heavily on training data. We also tested the structure prediction on 42 proteins from the TMB and compared to other existing algorithms. The results are comparable to existing algorithms, the accuracy ranges from 85-93%, depending upon the parameter used. This is very promising given that other algorithms rely heavily on homology and training datasets and may be overfitting. Our approach can be further improved by refining the energetic model, especially on turns and loops.

In addition, we have developed consensus methods to combine multiple secondary structures into one more reliable solution. Our results show that our technique can be used to combine multiple solutions to produce structures that are more than any of the input structures. These methods are based mainly on social choice theory and known properties of TMB proteins. In addition, we are working on methods for combining information on the super-secondary structures, and using them to augment the supersecondary structure provided by our approach.
5.3. Combinatorics and Annotation

5.3.1. Word counting and random generation

Cis-Regulatory modules (CRMs) of eukaryotic genes often contain multiple binding sites for transcription factors, or clusters. Formally, such sites can be viewed as *words* co-occurring in the DNA sequence. This gives rise to the problem of calculating the statistical significance of the event that multiple sites, recognized by different factors, would be found simultaneously in a text of a fixed length. A long-term research on word enumeration has been realized by the team. An extension to Hidden Markov Models has been realized recently in a collaboration with M. Roytberg (IMPB, Puschino, Russia). It relies on a new concept of overlap graphs that efficiently overcomes the main difficulty - overlapping occurrences - in probabilities computation. This is part of E. Furletova’s thesis, to be defended soon. An implementation is available at http://server2.lpm.org.ru/bio. This algorithm provides a significant space improvement over a previous algorithm, AHOPRO developed with our former associate team MIGEC. M. Régnier and S. Sheikh have addressed combinatorial problems on clumps that should allow further space decrease and large deviation results were presented at MCCMB’11.

An other application of word combinatorics has been started this year. During his internship, L. Pei (Paris-Sud 11 U.) provided a pipeline that simulates a random generation of reads and assembles them using MIRA software. This work will be pursued by D. Iakovishina in her thesis. It is a collaboration with MAGNOME at INRIA-BORDEAUX and IOGENE in Moscow.

A previous work [36], published in 2010, generalized Boltzmann samplers to multivariate objects, allowing for the efficient random generation achieving a fixed or approximate composition for context-free languages. However, the performances of such algorithms were only guaranteed in the case of strongly-connected context-free grammars. In a recent collaboration with O. Bodini, H. Tafat and C. Banderier (LIPN, Paris-XIII) we are working on characterizing the distributions arising from simply connected grammars. In a short paper accepted for presentation at the ANALCO’12 conference [24], we showed that: i) a large class of distributions can be reached for the number of occurrence of a single letter, arguably the simplest observable pattern; ii) simple grammars/regular expressions can be built that realize these distributions; iii) Classic Boltzmann samplers remain largely unaffected by this diversity.

Our work on random generation has applications in software testing and model-checking, in a collaboration with the Fortesse group at LRI [13], [29].

5.3.2. RNA combinatorics

Pseudoknots are usually ignored by popular software for RNA prediction. This means that, even under the daring assumptions of an unique and well-defined fold for RNA, coupled with a perfectly accurate energy model, the real structure of RNA will not be recovered perfectly. In a collaboration between AMIB members and S. Janssen (Universität Bielefeld), we investigated the practical implications of such a limitation. We used RNAFOLD, a popular software for the prediction of RNA structure on representative sequences of the RFAM database, which groups known RNA sequences into about 2000 functional families. We observed that 12% of RFAM families exhibited a total absence of overlap between predicted structures and manually-curated structures, derived from experimental or evolutionary data. Combination of RFAM annotations, a survey of literature, and a newly developed predictive method for the presence of a functional pseudoknots, we were able to validate that a large majority of the mispredicted families featured evidence of pseudoknots in the functional conformation. Preliminary results were presented by B. Raman at the Fifth Indo-French Bioinformatics Meeting [34].

In 2004, Condon and coauthors gave a hierarchical classification of exact RNA structure prediction algorithms according to the generality of structure classes that they handle. In [19], we completed this classification by adding two recent prediction algorithms. More importantly, we precisely quantified the hierarchy by giving closed or asymptotic formulas for the theoretical number of structures of given size $n$ in all the classes but one. This allows to assess the tradeoff between the expressiveness and the computational complexity of RNA structure prediction algorithms.
Similar decompositions can be used for the design of algorithms that include tractable subclasses of pseudo-knots. In [30] Y. Ponty and C. Saule extended a unifying framework introduced by Roytberg and Finkelstein to design ensemble RNA algorithms. This framework uses a family of hypergraphs to describe the conformation space, allowing for a clear separation between the search space, i.e. the set of admissible conformations, and the intended application (Minimal Free-Energy folding, partition function, statistical sampling...). We illustrated the promises of such an approach by explicitly rephrasing three major search spaces within the framework, and introduced an algorithm for computing the moments of any additive feature in the Boltzmann distribution.

By comparing empirical observations with the expected behavior of a model, combinatorial methods can be used to identify an evolutionary pressure weighing on RNA. In a collaboration with P. Clote (Boston College/DIGITEO) [11], we used analytic combinatorics to study the expected distance between both ends of an RNA molecule, or 5'-3' distance. Postulating a Boltzmann distribution on all secondary structures, we showed that this parameter is bounded by a – typically small – constant value when the sequence length goes to the infinity. Computing this quantity on a database of experimentally-determined secondary structures, we observed that the 5'-3' distances take larger values than those predicted from the model. Furthermore, quite surprisingly, this quantity was shown to correlate positively with the length. We concluded by hypothesizing that the secondary structure of RNA may be under evolutionary pressure to fold in a modular way, creating independent domains on the exterior face.

5.3.3. Data integration

Recent years have seen a revitalization of Data Integration research in the Life Sciences. But the perception of the problem has changed: While early approaches concentrated on handling schema-dependent queries over heterogeneous and distributed databases, current research emphasizes instances rather than schemas, tries to place the human back into the loop, and intertwines data integration and data analysis. In this domain, the contribution of AMIB in 2011 has been three folds: First, we have followed our collaboration with Ulf Leser (invited in the AMIB group at LR1 during 6 months in 2010) and have worked on the review of the past and current state of data integration for the Life Sciences and discussed recent trends in detail, which all pose various challenges for the database community in [28]. Additionally, we have worked on a vision of what should be done by workflow systems to make it possible to search, adapt, and reuse scientific workflows, the complete state-of-the-art on this domain has been provided [12]. Second, in close collaboration with oncologists from the Institut Curie and the Children’s Hospital of Philadelphia we have worked on the problem of ranking genes of interest associated to a given disease. The software GENEVALORIZATION has been designed and developed in this context and is able to provide a concise view of the literature available associated to a list of genes [10]. A second aspect of this research has been the design of a consensus ranking method, BioConsert, able to make the most (ie underline common points) of a set of established rankings [27]. This last point has been done in close collaboration with Sylvie Hamel invited professor in our group in 2010 (2 months). Third, we have presented a simple logical query language called RL for expressing different kinds of rules, especially well-suited to express association rules for transcriptomic data. In that context the challenge is to find out relationships between genes that reflect observations of how expression level of each gene affects those of others. The conjecture that association rules could be a model for the discovery of gene regulatory networks has already been partially validated. Nevertheless, several different kinds of rules between genes could be useful with respect to some biological objectives and we have designed a framework in which biologists may define their "own customized semantics" for rules with regard to their requirements. We have studied how the RL language behaves with respect to the well-known Armstrong’s axioms [22]. The main contribution of this paper is to exhibit a restricted form of RL-queries, yet with a good expressive power, for which Armstrong’s axioms are sound. From this result, this sublanguage turns out to have structural and computational properties which have been shown to be very useful in data mining, databases and formal concept analysis.

5.4. Systems Biology
In her thesis, M. Behzadi has developed a know-how on the behaviour of biological systems along a cooperation with an INSERM-INRA team based in Clermont-Ferrand. In the methodology that was developed, one computes the equations’ parameters from the experimental data in systems that can be considered at equilibrium. It was proved mathematically that some sub-domains are intrinsically stable and that their behaviour is not much affected by the initial conditions \[4\] for phospholipids biosynthesis. A review for carbone toxicity can also be found in \[15\]. Software Analyser software (MPSA) are currently under development by L. Paulevé.

Elementary flux mode is a fundamental concept as well as a useful tool in metabolic pathway analysis. However, when the networks are complex, the determination of elementary flux modes leads to combinatorial explosion of their number which prevents from drawing simple conclusions from their analysis. To deal with this problem, a biclustering method has developed \[18\] based on the Agglomeration of Common Motifs (ACoM). It was applied to the central carbon metabolism in Bacillus subtilis and to the yeast mitochondrial energy metabolism. It helped to give biological meaning to the different elementary flux modes and to the relatedness between reactions.

Once molecules and complexes participating in the signalling network have been identified, the relations between them (enzymatic reactions, activations, inhibitions, etc.) have to be deduced from experimental and literature data to build the influence graph. Partners INRA-BIOS and AMIB have started the development of a knowledge-based method, which uses the solver SOLAR, developed by NII (Tokyo), that allows automating this data integration task. We have already formalized the knowledge necessary for inferring the signalling network triggered by the FSH receptor, one famous GPCR. Preliminary results of this project ASAM are very encouraging.
6. New Results

6.1. Visual Analytics of EA Data

Participants: Jean-Daniel Fekete, Évelyne Lutton [correspondant].

An experimental analysis of Evolutionary Algorithms (EAs) usually generates a huge amount of multidimensional data, including numeric and symbolic data. It is difficult to efficiently navigate in such a set of data, for instance to be able to tune the parameters or evaluate the efficiency of some operators. Usual features of existing EA visualisation systems consist in visualising time- or generation-dependent curves (fitness, diversity, or other statistics). When dealing with genomic information, the task becomes even more difficult, as a convenient visualisation strongly depends on the considered fitness landscape. In this latter case the raw data are usually sets of successive populations of points of a complex multidimensional space.

The purpose of this study was to evaluate GraphDice on complex sets of EA data (for artificial and real test-cases), and to sketch future developments of this tool, in order to better adapt it to the needs of EA experimental analysis (Fig. 6). An output of this study is the acceptance of the EASEA-Cloud ANR-Emergence project, in which developments will aim at adding tools in GraphDice specific for:

- visualisation of Evolutionary Algorithms written in the EASEA language,
- monitoring the execution of these algorithms on a cloud of computers (CPU + GPU).
6.2. Interactive Evolutionary Algorithms for Visual decision making

**Participants:** Nadia Boukhelifa, Waldo Cancino, Jean-Daniel Fekete, Evelyne Lutton [correspondant].

When dealing with very large datasets with many dimensions, it is often difficult to efficiently navigate and find interesting viewpoints, signficative compound variables, unexpected behaviour, and other remarkable characteristics.

Our aim within the Systematic CSDL project (Complex Systems Design Lab, 2009–2012) is to use interative evolutionary algorithm to assist the user in its exploration task. Finding an interesting, non obvious, viewpoint on a complex dataset can be formulated as an interactive optimisation problem. Population-based evolutionary search mechanisms can then efficiently be exploited for suggesting new viewpoints on data, that progressively adapt to the needs of an user.

In September 2011 (arrival of Nadia Boukhelifa and Waldo Cancino) we started to build a prototype based on GraphDice, that proposes new dimensions in the scatterplot matrix. These secondary set of dimensions are compositions of the dimensions of the initial datased. Starting from an initial set of suggested dimensions (PCA analysis of the dataset), an evolutionary algorithm progressively refines the compound dimensions according to a measurement of the activity of the user on the corresponding views.

6.3. Optimisation of Food Models

**Participant:** Évelyne Lutton [correspondant].

_In collaboration with Alberto Tonda and Romain Reuillon, ISC-PIF_

The European project DREAM (http://dream.csregistry.org/) managed by INRA-CEPIA, aims at building decision support tools for better managing product quality and, by the way, manufacturing processes in the domain of agrifood industry.

Our contribution to this project is focused on the evolutionary optimisation of Bayesian Networks models, on the development of efficient cooperative-co-evolution schemes to solve some food modeling problems (milk gel, cheese ripening), and on the efficient visualisation of output data of these algorithms.

6.4. A Study on Dual-Scale Data Charts

We presented the results of a user study that compares different ways of representing dual-scale data charts (see Fig. 7). Dual-scale charts incorporate two different data resolutions into one chart in order to emphasize data in regions of interest or to enable the comparison of data from distant regions. While some design guidelines exist for these types of charts, there is currently little empirical evidence on which to base their design. We filled this gap by discussing the design space of dual-scale cartesian-coordinate charts and by experimentally comparing the performance of different chart types with respect to elementary graphical perception tasks such as comparing lengths and distances. Our study suggests that cut-out charts which include collocated full context and focus are the best alternative, and that superimposed charts in which focus and context overlap on top of each other should be avoided.

6.5. Information Visualization Evaluation

Petra Isenberg has contributed to three articles on evaluation methodologies: The first article “Collaborative Visualization: Definition, Challenges, and Research Agenda” [12] deals with challenges of collaborative visual analytics and includes a discussion on the challenges of evaluating tools during multi-person use. The second article, “Information Visualization Evaluation in Large Companies: Challenges, Experiences and Recommendations” [14], discusses challenges of evaluating and deploying visual analytics tools in a large company setting. It lists several challenges and provides concrete guidance to others who seek to evaluate tool within domain experts in their work environment. Finally, “Seven Guiding Scenarios for Information Visualization Evaluation” [33] is a pre-print of an journal article (in press) which provides a new viewpoint on evaluation in information visualization. Instead of giving an overview of methods, it cites evaluation goals and questions and can, thus, provide clear considerations for practitioners and researchers in the area.
4. New Results

4.1. Iterative Optimization for the Data Center (Alchemy-related research)

This result corresponds to research started within Alchemy, and it is less related to the objectives of ByMoore itself.

Iterative optimization is a simple but powerful approach that searches for the best possible combination of compiler optimizations for a given workload. However, each program, if not each data set, potentially favors a different combination. As a result, iterative optimization is plagued by several practical issues that prevent it from being widely used in practice: a large number of runs are required for finding the best combination; the process is inherently data set sensitive; and the exploration process incurs significant overhead that needs to be compensated for by performance benefits. Therefore, while iterative optimization has been shown to have significant performance potential, it is seldomly used in production compilers.

We propose [4] Iterative Optimization for the Data Center (IODC): we show that servers and data centers offer a context in which all of the above hurdles can be overcome. The basic idea is to spawn different combinations across workers and recollect performance statistics at the master, which then evolves to the optimum combination of compiler optimizations. IODC carefully manages costs and benefits, and is transparent to the end user.

We evaluate IODC using both MapReduce and throughput server applications. In order to reflect the large number of users interacting with the system, we gather a very large collection of data sets (at least 1000 and up to several million unique data sets per program), for a total storage of 10.7TB, and 568 days of CPU time. We report an average performance improvement of $1.48 \times$, and up to $2.08 \times$, for the MapReduce applications, and $1.14 \times$, and up to $1.39 \times$, for the throughput server applications.

4.2. Statistical Performance Comparisons of Computers (Alchemy-related research)

This result corresponds to research started within Alchemy, and it is less related to the objectives of ByMoore itself.

As a fundamental task in computer architecture research, performance comparison has been continuously hampered by the variability of computer performance. In traditional performance comparisons, the impact of performance variability is usually ignored (i.e., the means of performance measurements are compared regardless of the variability), or in the few cases where it is factored in using parametric confidence techniques, the confidence is either erroneously computed based on the distribution of performance measurements, instead of the distribution of sample mean of performance measurements, or too few measurements are considered for the distribution of sample mean to be normal. We first illustrate how such erroneous practices can lead to incorrect comparisons.

Then, we propose [3] a non-parametric Hierarchical Performance Testing (HPT) framework for performance comparison, which is significantly more practical than standard parametric confidence tests because it does not require to collect a large number of measurements in order to achieve a normal distribution of the sample mean. This HPT framework has been implemented as an open-source software.
4.3. Implementation of Signal Processing Tasks on Neuromorphic Hardware

Because of power and reliability issues, computer architects are forced to explore new types of architectures, such as heterogeneous systems embedding hardware accelerators. Neuromorphic systems are good candidate accelerators that can perform efficient and robust computing for certain classes of applications. We propose\[ 9 \] a spiking neurons based accelerator, with its hardware and software, that can be easily programmed to execute a wide range of signal processing applications. A library of operators is built to facilitate implementation of various types of applications. Automated placement and routing software tools are used to map these applications onto the hardware. Altogether, this system aims at providing to the user a simple way to implement signal processing tasks on neuromorphic hardware.

4.4. Automatic Abstraction and Fault Tolerance in Cortical Microarchitectures

Recent advances in the neuroscientific understanding of the brain are bringing about a tantalizing opportunity for building synthetic machines that perform computation in ways that differ radically from traditional Von Neumann machines. These brain-like architectures, which are premised on our understanding of how the human neocortex computes, are highly fault-tolerant, averaging results over large numbers of potentially faulty components, yet manage to solve very difficult problems more reliably than traditional algorithms. A key principle of operation for these architectures is that of automatic abstraction: independent features are extracted from highly disordered inputs and are used to create abstract invariant representations of the external entities. This feature extraction is applied hierarchically, leading to increasing levels of abstraction at higher levels in the hierarchy. This work \[ 6 \] describes and evaluates a biologically plausible computational model for this process, and highlights the inherent fault tolerance of the biologically-inspired algorithm. We introduce a stuck-at fault model for such cortical networks, and describe how this model maps to hardware faults that can occur on commodity GPGPU cores that used to realize the model in software. We show experimentally that the model software implementation can intrinsically preserve its functionality in the presence of faulty hardware, without requiring any reprogramming or recompilation. This model is a first step towards developing a comprehensive and biologically-plausible understanding of the computational algorithms and microarchitecture of computing systems that mimic the human cortex, and to applying them to the robust implementation of computational tasks on future computing systems built of faulty components.
6. New Results

6.1. Foundations of information hiding

Information hiding refers to the problem of protecting private information while performing certain tasks or interactions, and trying to avoid that an adversary can infer such information.

This is one of the main areas of research in Comète, and two PhD thesis based on this topic have been defended this year in Comète [12], [11] have been defended this year. We are exploring several topics, described below. An overview of our results is contained in [24].

6.1.1. The problem of information hiding in presence of concurrency

The analysis of probabilistic concurrent systems usually relies on the notion of scheduler in order to solve the nondeterminism. Unfortunately the classical notion of scheduler, which is a mathematical functions that chooses the next step depending on the history of the computation, can leak any secret information contained in the history. This creates false positives, and it is known as the problem of the allmighty scheduler. One way to solve this problem, already explored in literature, is to fix the strategy of the scheduler beforehand [31]. However this solution is considered too rigid and unrealistic. In [14] we have propose a milder restriction on the schedulers, and we have defined the notion of strong (probabilistic) information hiding under various notions of observables. Furthermore, we have proposed a method, based on the notion of automorphism, to verify that a system satisfies the property of strong information hiding, namely strong anonymity or no-interference, depending on the context.

6.1.2. Modeling the knowledge of the adversary

In [15] we have developed a game semantics for process algebra with two interacting agents. The purpose of our semantics is to make manifest the role of knowledge and information flow in the interactions between agents and to control the information available to interacting agents. We have defined games and strategies on process algebras, so that two agents interacting according to their strategies determine the execution of the process, replacing the traditional scheduler. We have shown that different restrictions on strategies represent different amounts of information being available to a scheduler. We have also shown that a certain class of strategies corresponds to the syntactic schedulers of Chatzikokolakis and Palamidessi [32], which were developed to overcome problems with traditional schedulers modeling interaction. The restrictions on these strategies have an explicit epistemic flavor.

6.1.3. Opacity

Opacity is a security property formalizing the absence of secret information leakage and we have addressed in [30] the problem of synthesizing opaque systems. A secret predicate \( S \) over the runs of a system \( G \) is opaque to an external user having partial observability over \( G \), if s/he can never infer from the observation of a run of \( G \) that the run belongs to \( S \). We have chosen to control the observability of events by adding a device, called a mask, between the system \( G \) and the users. We have first investigated the case of static partial observability where the set of events the user can observe is fixed a priori by a static mask. In this context, we have shown that checking whether a system is opaque is PSPACE-complete, which implies that computing an optimal static mask ensuring opacity is also a PSPACE-complete problem. Then, we have introduced dynamic partial observability where the set of events the user can observe changes over time and is chosen by a dynamic mask. We have shown how to check that a system is opaque with respect to a dynamic mask and we have also addressed the corresponding synthesis problem: given a system \( G \) and secret states \( S \), compute the set of dynamic masks under which \( S \) is opaque. Our main result is that the set of such masks can be finitely represented and can be computed in EXPTIME and this is a lower bound. Finally we have also addressed the problem of computing an optimal mask.
6.1.4. Interactive systems

In [13] we have considered systems where secrets and observables can alternate during the computation. We have shown that the information-theoretic approach which interprets such systems as (simple) noisy channels is not valid anymore. However, the principle can be recovered if we consider more complicated types of channels, that in Information Theory are known as channels with memory and feedback. We have shown that there is a complete correspondence between interactive systems and such kind of channels. Furthermore, we have shown that the capacity of the channels associated to such systems is a continuous function of the Kantorovich metric.

6.1.5. Differential privacy

Differential privacy is a notion that has emerged in the community of statistical databases, as a response to the problem of protecting the privacy of the database’s participants when performing statistical queries. The idea is that a randomized query satisfies differential privacy if the likelihood of obtaining a certain answer for a database $x$ is not too different from the likelihood of obtaining the same answer on adjacent databases, i.e. databases which differ from $x$ for only one individual.

In [17], [16], we have analyzed critically the notion of differential privacy in light of the conceptual framework provided by the Rényi min information theory. We have shown that there is a close relation between differential privacy and leakage, due to the graph symmetries induced by the adjacency relation. Furthermore, we have considered the utility of the randomized answer, which measures its expected degree of accuracy. We have focused on certain kinds of utility functions called “binary”, which have a close correspondence with the Rényi min mutual information. Again, it turns out that there can be a tight correspondence between differential privacy and utility, depending on the symmetries induced by the adjacency relation and by the query. Depending on these symmetries we can also build an optimal-utility randomization mechanism while preserving the required level of differential privacy. Our main contribution is a study of the kind of structures that can be induced by the adjacency relation and the query, and how to use them to derive bounds on the leakage and achieve the optimal utility.

6.2. Concurrent constraint programming

6.2.1. Bisimilarity

Bisimilarity is one of the main representative equivalences for concurrent behaviour. It captures our intuitive notion of process equivalence; two processes are equivalent if they can match each other's moves. Furthermore, it provides an elegant co-inductive proof technique based on the notion of bisimulation. Nevertheless, there have been few attempts to define a notion of bisimilarity for concurrent constraint programming (ccp). The ones we were aware of are those in [40] and [36] but they are not completely satisfactory: The first one may tell apart processes with identical observable behaviour, while the second quantifies over all possible inputs from the environment, and hence it is not clear whether it can lead to a feasible proof technique.

Bisimilarity relies on labelled transitions: each evolution step of a system is tagged by some information aimed at capturing the possible interactions of a process with the environment. In [18] we have provided a labelled transition system for ccp and we have proposed a notion of ccp bisimilarity. Intuitively, in this transition system the labels represent the minimal information that processes require from the environment to execute. Furthermore we have shown that, unlike previous approaches, our notion of bisimilarity coincides with the standard notion of equivalence for (deterministic) ccp. This way we have provided ccp with an alternative co-inductive proof technique, coherent with previous equivalences, for process behaviour.

When the state space of a system is finite, the ordinary notion of bisimilarity can be computed via the well-known partition refinement algorithm, but unfortunately, this algorithm does not work for ccp bisimilarity. In [19] we have proposed a variation of the partition refinement algorithm for verifying ccp bisimilarity. To the best of our knowledge this is the first work providing for the automatic verification of program equivalence for ccp.
6.2.2. Modeling cellular signaling systems

The molecular mechanisms of cell communication with the environment involve many concurrent processes governing dynamically the cell function. This concurrent behavior makes traditional methods, such as differential equations, unsatisfactory as a modeling strategy since they do not scale well when a more detailed view of the system is required.

In [19] we have described a modeling strategy for cellular signaling systems based on a temporal and probabilistic extension of CCP. Starting from an abstract model, we have built refinements adding further details coming from experimentation or abstract assumptions. The advantages of our approach are: due to the notion of partial information as constraints in CCP, the model can be straightforwardly extended when more information is available; qualitative and quantitative information can be represented by means of probabilistic constructs of the language; finally, the model is a runnable specification and can be executed, thus allowing for the simulation of the system. We have outlined the use of this methodology to model the interaction of G-protein-coupled receptors with their respective G-proteins that activates signaling pathways inside the cell. Finally, we have presented simulation results obtained from an implementation of the framework.

6.3. Session types

In [22] we have presented a type checking algorithm for establishing a session-based discipline in the pi calculus of Milner, Parrow and Walker [37]. Our session types are qualified as linear or unrestricted. Linearly typed communication channels are guaranteed to occur in exactly one thread, possibly multiple times; afterwards they evolve as unrestricted channels. Session protocols are described by a type constructor that denotes the two ends of one and the same communication channel. We have proved the soundness of the algorithm by showing that processes consuming all linear resources are accepted by a typing system preserving typings during the computation and that type checking is consistent w.r.t. structural congruence.
6. New Results

6.1. Optimal control with singular arcs

Participants: Pierre Martinon, Andrei Dmitruk [Moscow State University], Pablo Lotito [U. Tandil, Argentina], Soledad Aronna, Frédéric Bonnans.

These studies enter in the framework of the PhD thesis of S. Aronna, supervised by J.F. Bonnans and P. Lotito, that ended in December 2011.

In the paper [21] we deal with optimal control problems for systems affine in the control variable. We have nonnegativity constraints on the control, and finitely many equality and inequality constraints on the final state. First, we obtain second order necessary optimality conditions. Secondly, we get a second order sufficient condition for the scalar control case. The results use in an essential way the Goh transformation.

In the report [22], we design a shooting algorithm applied to optimal control problems for which all control variables enter linearly in the Hamiltonian. This shooting algorithm is non standard, in particular since there are more equations than unknowns, and extends some previous algorithms designed for specific structures. We start investigating the case having only initial-final state constraints and free control variable, and afterwards we deal with control bounds. The shooting algorithm is locally well-posed and quadratically convergent if the derivative of its associated shooting function is injective at the optimal solution. The main result of this paper is to provide a sufficient condition for this injectivity, that is very close to the second order necessary condition. We prove that this sufficient condition guarantees the stability of the optimal solution under small perturbations and the well-posedness of the shooting algorithm for the perturbed problem. We present numerical tests that validate our method.

In the report [20] we deal with optimal control problems for systems that are affine in one part of the control variables and nonlinear in the rest of the control variables. We have finitely many equality and inequality constraints on the initial and final states. First we obtain second order necessary and sufficient conditions for weak optimality. Afterwards, we propose a shooting algorithm, and we show that the sufficient condition above-mentioned is also sufficient for the injectivity of the shooting function at the solution.

6.2. Characterization of a local quadratic growth of the Hamiltonian for control constrained optimal control problems

Participants: Frédéric Bonnans, Nikolai Osmolovskii [Systems Research Institute, Warsaw].

In the paper [25] we consider an optimal control problem with inequality control constraints given by smooth functions satisfying the hypothesis of linear independence of gradients of active constraints. For this problem, we formulate a generalization of strengthened Legendre condition and prove that this generalization is equivalent to the condition of a local quadratic growth of the Hamiltonian subject to control constraints.

6.3. Hamilton-Jacobi approach for deterministic control problems

Participants: Albert Altarovici, Olivier Bokanowski, Yingda Cheng [University of Texas], Anna Desilles, Nicolas Forcadel, Zhiping Rao, Chi-Wang Shu [Brown University], Hasnaa Zidani.
The paper [30] deals with deterministic optimal control problem with state constraints and non-linear dynamics. It is known for such a problem that the value function is in general discontinuous and its characterization by means of an HJ equation requires some controllability assumptions involving the dynamics and the set of state constraints. Here, we first adopt the viability point of view and look at the value function as its epigraph. Then, we prove that this epigraph can always be described by an auxiliary optimal control problem free of state constraints, and for which the value function is Lipschitz continuous and can be characterized, without any additional assumptions, as the unique viscosity solution of a Hamilton-Jacobi equation. The idea introduced in this paper bypass the regularity issues on the value function of the constrained control problem and leads to a constructive way to compute its epigraph by a large panel of numerical schemes. Our approach can be extended to more general control problems. We study in this paper the extension to the infinite horizon problem as well as for the two-player game setting. Finally, an illustrative numerical example is given to show the relevance of the approach.

In [34], [19] we study an optimal control problem governed by measure driven differential systems and in presence of state constraints. First, under some weak invariance assumptions, we study in [19] the properties of the value function and obtain its characterization by means of an auxiliary control problem of absolutely continuous trajectories. For this, we use some known techniques of reparametrization and graph completion. Then we give a characterization of the value function as the unique constrained viscosity solution of a Hamilton-Jacobi equation with measurable time dependent Hamiltonians.

The general case without assuming any controllability assumption is considered in [34]. We prove that the optimal solutions can still be obtained by solving a reparametrized control problem of absolutely continuous trajectories but with time-dependent state-constraints.

The paper [17] deals with minimal time problems governed by nonlinear systems under general time dependent state constraints and in the two-player games setting. In general, it is known that the characterization of the minimal time function, as well as the study of its regularity properties, is a difficult task in particular when no controllability assumption is made. In addition to these difficulties, we are interested here to the case when the target, the state constraints and the dynamics are allowed to be time-dependent. We introduce a particular reachability control problem, which has a supremum cost function but is free of state constraints. This auxiliary control problem allows to characterize easily the backward reachable sets, and then, the minimal time function, without assuming any controllability assumption. These techniques are linked to the well known level-set approaches. Our results can be used to deal with motion planning problems with obstacle avoidance, see [16].

Several works have been also carried out in the domain of numerical methods of HJB equations. The paper [31] aims at studying a discontinuous Galerkin scheme for front propagation with obstacles. We extend a first work published in [11], to propose a simple and direct discontinuous Galerkin (DG) method adapted to such front propagation problems. We follow the formulation of [12], leading to a level set formulation driven by a Hamilton-Jacobi variational inequality. The DG scheme is motivated by the variational formulation when the equation corresponds to linear convection problems in presence of obstacles. The scheme is then generalized to nonlinear equations, written in an explicit form. Stability analysis are performed for the linear case with Euler forward, a Heun scheme and a Runge-Kutta third order time discretization. Several numerical examples are provided to demonstrate the robustness of the method. Finally, a narrow band approach is considered in order to reduce the computational cost.

6.4. Stochastic programming

Participants: Frédéric Bonnans, Zhihao Cen, Thibault Christel [Total].

In [29] we consider a model of medium-term commodity contracts management. Randomness takes place only in the prices on which the commodities are exchanged whilst state variable is multi-dimensional. In our previous article, we proposed an algorithm to deal with such problem, based on quantization of random process and a dual dynamic programming type approach. We obtained accurate estimates of the optimal value and a suboptimal strategy from this algorithm. In this paper, we analyse the sensitivity with respect to
parameters driving the price model. We discuss the estimate of marginal price based on the Danskin’s theorem. Finally, some numerical results applied to realistic energy market problems have been performed. Comparisons between results obtained by our algorithm and other classical methods are provided and evidence the accuracy of the estimate of marginal prices.

6.5. Stochastic control

**Participants:** Frédéric Bonnans, Xiaolu Tan [CMAP], Imene Ben Latifa, Mohamed Mnif [ENIT, Tunis].

In [24], we extend a study by Carmona and Touzi on an optimal multiple stopping time problem in a market where the price process is continuous. In this paper, we generalize their results when the price process is allowed to jump. Also, we generalize the problem associated to the valuation of swing options to the context of jump diffusion processes. Then we relate our problem to a sequence of ordinary stopping time problems. We characterize the value function of each ordinary stopping time problem as the unique viscosity solution of the associated Hamilton-Jacobi-Bellman Variational Inequality.

In [27], we consider, in the framework of Galichon, Henry-Labordère and Touzi, the model-free no-arbitrage bound of variance option given the marginal distributions of the underlying asset. We first make some approximations which restrict the computation on a bounded domain. Then we propose a gradient projection algorithm together with a finite difference scheme to approximate the bound. The general convergence result is obtained. We also provide a numerical example on the variance swap option.

6.6. Stochastic control of an hybrid vehicle

**Participants:** Kamal Aouchiche [Renault], Frédéric Bonnans, Giovanni Granato, Hasnaa Zidani.

In the CDC paper [18] we present a stochastic dynamic programming (SDP) algorithm that aims at minimizing an economic criteria based on the We also work on a stochastic dynamic programming (SDP) algorithm that aims at minimizing an economic criteria based on the total energy consumption of a range extender electric vehicle (REEV). This algorithm integrates information from the REEV’s navigation system in order to obtain some information about future expected vehicle speed. The model of the vehicle’s energetic system, which consists of a high-voltage (HV) battery, the main energy source, and an internal combustion engine (ICE), working as an auxiliary energy source), is written as a hybrid dynamical system and the associated optimization problem in the hybrid optimal control framework. The hybrid optimal control problem includes two important physical constraints on the ICE, namely, an activation delay and a decision lag. Three methods for the inclusion of such physical constraints are studied. After introducing the SDP algorithm formulation we comment on numerical results of the stochastic algorithm and its deterministic counterpart.

6.7. Optimal control of PDEs

**Participants:** Frédéric Bonnans, Francisco Silva [U. Roma], Térence Bayen [U. Montpellier II].

In the report [23] we consider an optimal control problem of a semi-linear elliptic equation, with bound constraints on the control. Our aim is to characterize local quadratic growth for the cost function $J$ in the sense of strong solutions. This means that the function $J$ grows quadratically over all feasible controls whose associated state is close enough to the nominal one, in the uniform topology. The study of strong solutions, classical in the Calculus of Variations, seems to be new in the context of PDE optimization. Our analysis, based on a decomposition result for the variation of the cost, combines Pontryagin’s principle and second order conditions. While these two ingredients are known, we use them in such a way that we do not need to assume that the Hessian of Lagrangian of the problem is a Legendre form, or that it is uniformly positive on an extended set of critical directions.

6.8. Global optimization of pipe networks by the interval analysis approach: the Belgium network case

**Participants:** Frédéric Bonnans, Grégoire Spiers, Jean-Léopold Vie.
In [26] we have shown that a classical test problem for the optimization of gas networks, namely the so-called Belgium gas network, could be solved by global optimization techniques. Until now only local algorithms had been used for solving this problem. Using techniques based on interval analysis and constraint propagation we actually recover (and therefore justify) the solution computed so far.
5. New Results

5.1. Specification and verification of database driven systems

**Participants:** Serge Abiteboul, Pierre Bourhis, Luc Segoufin, Szymon Toruńczyk, Victor Vianu.

Modelization and verification of data centric systems. We have intensively studied the Active XML model. It is a high-level specification language tailored to data-intensive, distributed, dynamic Web services. Active XML is based on XML documents with embedded function calls. The state of a document evolves depending on the result of internal function calls (local computations) or external ones (interactions with users or other services). Function calls return documents that may be active, so may activate new subtasks. Our first line of result is a comparison of the specification power of various workflow control mechanisms within the Active XML framework and beyond [23].

AXML is very powerful and many static analysis problems are undecidable. We have also introduced a model of automata designed for modeling infinite runs of systems equipped with static relational databases. The automata model is equipped with finitely many variables, each of which can store values from a linearly ordered domain, such as the natural numbers. The transitions of the automata depends on a conjunctive query involving the database and the current values of the variables. For verifying infinite runs of such automata, an extension of temporal logic is considered, capable of comparing values stored in the variables and the database, at different times of the run. The main contribution of the work is the proof that automated verification of such temporal properties of the system can be carried out in PSpace. For more details, see [35].

Static analysis of query languages. XPath is arguably the most widely used XML query language as it is implemented in XSLT and XQuery and it is used as a constituent part of several specification and update languages. Hence in order to perform static analysis on a system manipulating XML data it is important to master the static analysis for XPath. Most of the important static analysis problems reduce to satisfiability checking: does a given query return a non-empty answer on some data. In general, in the presence of data values, the satisfiability of XPath is undecidable. We have shown that when restricted to its vertical navigational power, XPath becomes decidable [30].

5.2. Distributed data management

**Participants:** Serge Abiteboul, Emilien Antoine, Daniel Deutch, Alban Galland, Wojciech Kazana, Yannis Katsis, Luc Segoufin, Cristina Sirangelo.

Distributed knowledge base. As a foundation for managing distribution, we have proposed a model of a distributed knowledge base, that handles data and meta-data, as well as access control and localization, in a unique integrated setting. To support automatic reasoning on this knowledge base, we also introduced a novel rule-based language supporting the exchange of rules, namely Webdamlog. This work has been presented [21] and demonstrated [26] at major database conferences.

Probabilistic XML. Data from the Web are imprecise and uncertain. To manage this imprecision in a well-principled way, we have made significant advances in the field of probabilistic databases, and specifically, probabilistic XML. We have introduced new tractable probabilistic models for representing uncertain hierarchical information, and carried out in-depth studies of query evaluation, aggregation, and updates in various probabilistic XML models. These results have matured and some of the results are available in journal articles, e.g., [14].
Enumeration of query answers. In many applications the output of a query may have a huge size and enumerating all the answers may already consume too many of the allowed resources. In this case it may be appropriate to first output a small subset of the answers and then, on demand, output a subsequent small numbers of answers and so on until all possible answers have been exhausted. To make this even more attractive it is preferable to be able to minimize the time necessary to output the first answers and, from a given set of answers, also minimize the time necessary to output the next set of answers - this second time interval is known as the delay. We have shown that this was doable with a linear preprocessing time and constant enumeration delay for first-order queries over structures of bounded degree [19].

Data exchange and Web incomplete information. We have addressed the problem of restructuring data exchanged between communicating applications on the Web. We have proposed and analyzed a new language to specify data restructuring rules (schema mappings). This language generalizes existing mapping dependencies, by allowing a more flexible specification mechanism [20].

Jorge. We also invested a lot of effort in a textbook (undergraduate and graduate level) on Web data management (nicknamed Jorge) to be published at Cambridge University Press [38]. The book is already available on the Webdam Web site http://webdam.inria.fr/Jorge

5.3. Tree automata theory

Participants: Stéphane Demri, Florent Jacquemard, Luc Segoufin.

Most of our results for this section concerns data words and data trees. Those are words and trees where each position contains a data value together with the classical label. Data trees can be seen as a model for XML data. We have studied automata model using registers or memory or allowing constraints that can involve data comparisons in its transitions.

Register Automata. These extend the classical model of finite automata with auxiliary registers storing data values for later comparison.

We have introduced a new model of automata over data trees and shown the decidability of its emptiness problem [30]. These automata were used for obtaining decidability results for the static analysis for some fragments of XPath presented in the previous section.

Automata with counters. In [39], a survey chapter on the verification of infinite-state systems is presented that is focused on the verification of counter systems. Verification problems for vector addition systems and recursive Petri nets are considered. Moreover, we introduce subclasses of counter systems for which reachability questions can be solved in Presburger arithmetic viewed as a means to symbolically represent sets of tuples of natural numbers.

Automata with isomorphism tests among subtrees. We have also considered some models described by tree automata enriched with a feature testing for isomorphisms between subtrees. Such constraints could be used for testing monadic key constraints over XML documents. For these models, the main challenge is to establish the decidability of the non-emptiness of the language specified by a given automaton [18].

Rewriting Controlled by Selection Automata. Motivated by the problem of static analysis of XML update programs, we have studied [33] the combination, called controlled term rewriting systems (CTRS), of term rewriting rules with constraints selecting the possible rewrite positions. These constraints are specified, for each rewrite rule, by a selection automaton which defines a set of positions in a term based on tree automata computations. We have established several decidability and complexity results for several cases of the reachability and regular model checking problems for this tree transformation formalism.
6. New Results

6.1. Sampling methods for inverse scattering problems

6.1.1. Sampling methods with time dependent data

**Participants:** Houssem Haddar, Armin Lechleiter, Simon Marmorat.

We considered the extension of the so-called Factorization method to far-field data in the time domain. For a Dirichlet scattering object and incident wave fronts, the inverse problem under investigation consists in characterizing the shape of the scattering object from the behaviour of the scattered field far from the obstacle (far-field measurements). We derive a self-adjoint factorization of the time-domain far-field operator and show that the middle operator of this factorization possesses a weak type of coercivity. This allows to prove range inclusions between the far-field operator and the time-domain Herglotz operator.

We also extended the near-field version of the linear sampling method to causal time-dependent wave data for smooth, band-limited incident pulses, considering different boundary conditions as for instance Dirichlet, Neumann or Robin conditions [27].

6.1.2. Inverse problems for periodic penetrable media

**Participants:** Armin Lechleiter, Dinh Liem Nguyen.

Imaging periodic penetrable scattering objects is of interest for non-destructive testing of photonic devices. The problem is motivated by the decreasing size of periodic structures in photonic devices, together with an increasing demand in fast non-destructive testing. In this project, linked to the thesis project of Dinh Liem Nguyen, we considered the problem of imaging a periodic penetrable structure from measurements of scattered electromagnetic waves. As a continuation of earlier work, we considered an electromagnetic problem for transverse magnetic waves (earlier work treats transverse electric fields), and also the full Maxwell equations. In both cases, we treat the direct problem by a volumetric integral equation approach and construct a Factorization method.

6.1.3. Inverse problems for Stokes-Brinkmann flows

**Participants:** Armin Lechleiter, Tobias Rienmüller.

Geometric inverse problems for flows arise for instance when controlling pipelines and oil reservoirs. In this project, we considered the Stokes-Brinkmann equations that model, for instance, porous penetrable inclusions in a free background. The factorization method is able to characterize the inclusions from the relative Dirichlet-to-Neumann operator. Numerical examples show the feasibility of the method.

6.1.4. Inverse scattering from screens with impedance boundary conditions

**Participants:** Yosra Boukari, Houssem Haddar.

We are interested in solving the inverse problem of determining a screen (or a crack) from multi-static measurements of electromagnetic (or acoustic) scattered field at a given frequency. An impedance boundary condition is assumed to be verified at both faces of the screen. We extended the so-called factorization method to this setting. We also analyzed a data completion algorithm based on integral equation method for the Helmholtz equation. This algorithm is then coupled to the so-called RG-LSM algorithm to retrieve cracks inside a locally homogeneous background. This work is conducted in collaboration with F. Ben Hassen.

6.1.5. Transmission Eigenvalues and their application to the identification problem

**Participants:** Anne Cossonnière, Houssem Haddar, Giovanni Giorgi.
The so-called interior transmission problem plays an important role in the study of inverse scattering problems from (anisotropic) inhomogeneities. Solutions to this problem associated with singular sources can be used for instance to establish uniqueness for the imaging of anisotropic inclusions from multi-static data at a fixed frequency. It is also well known that the injectivity of the far field operator used in sampling methods is equivalent to the uniqueness of solutions to this problem. The frequencies for which this uniqueness fails are called transmission eigenvalues. We are currently developing approaches where these frequencies can be used in identifying (qualitative informations on) the medium properties. Our research on this topic is mainly done in the framework of the associate team ISIP http://www-direction.inria.fr/international/PHP/Networks/LiEA.php with the University of Delaware. A review article on the state of art concerning the transmission eigenvalue problem has been written in collaboration with F. Cakoni [24].

The main topic of the PhD thesis of A. Cossonnière is to extend some of the results obtained above (for the scalar problem) to the Maxwell’s problem. In this perspective, theoretical results related to solutions of the interior transmission problem for medium with cavities and existence of transmission eigenvalues have been obtained [14]. This work is then extended to the case of medium with perfectly conducting inclusions. Only the scalar case has been studied [35]. In collaboration with M. Fares and F. Collino from CERFACS we investigated the use of a surface integral equation approach to find the transmission eigenvalues for inclusions with piecewise constant index. The main difficulty behind this procedure is the compactness of the obtained integral operator in usual Sobolev spaces associated with the forward scattering problem. We solved this difficulty by introducing a preconditioning operator associated with a “coercive” transmission problem. The obtained procedure has been validated numerically in 2D and 3D cases. We also analyzed the transmission eigenvalue problem using this surface integral equation approach. This technique allowed us to generalize discretness results on the spectrum to cases where the contrast can change sign [2].

With G. Giorgi, we developed a method that give estimates on the material properties using the first transmission eigenvalue. This method is based on reformulating the interior transmission eigenvalue problem into an eigenvalue problem for the material coefficients. We validated our methodology for homogeneous and inhomogeneous inclusions and backgrounds. We also treated the case of a background with absorption and the case of scatterers with multiple connected components of different refractive indexes [26].

6.1.6. The factorization method for EIT with inhomogeneous background

Participants: Giovanni Migliorati, Houssem Haddar.

We developed a numerical inversion scheme based on the Factorization Method to solve the (continuous model of) Electrical Impedance Tomography problem with inhomogeneous background. The numerical scheme relies on the well chosen approximation by the finite element method of the solution to the dipole-like Neumann boundary-value problem. Two regularization techniques are tested, i.e. the Tikhonov regularization embedding Morozov principle, and the classical Picard Criterion. The numerical analysis of the method and the results obtained are presented in the INRIA report [28].

6.2. Iterative Methods for Non-linear Inverse Problems

6.2.1. Inverse medium problem for axisymmetric eddy current models

Participants: Houssem Haddar, Zixian Jiang, Armin Lechleiter.

We are interested in shape optimization methods for inclusion detection in an axisymmetric eddy current model. This problem is motivated by non-destructive testing methodologies for steam generators. We investigated the validity of the eddy current model for these kinds of problems and developed numerical methods for the solution of the direct problem in weighted Sobolev spaces. Then we computed the shape derivative of an inclusion which allows to use regularized iterative methods to solve the inverse problem [23]. We also develop asymptotic models to identify thin highly conducting deposits.

6.2.2. Hybrid methods for inverse scattering problems

Participants: Grégoire Allaire, Houssem Haddar, Dimitri Nicolas.
It is well admitted that optimization methods offer in general a good accuracy but are penalized by the cost of solving the direct problem and by requiring a large number of iterations due to the ill-posedness of the inverse problem. However, profiting from good initial guess provided by sampling methods these method would become viable. Among optimization methods, the Level Set method seems to be well suited for such coupling since it is based on capturing the support of the inclusion through an indicator function computed on a cartesian grid of probed media. Beyond the choice of an optimization method, our goal would be to develop coupling strategies that uses sampling methods not only as an initialization step but also as a method to optimize the choice of the incident (focusing) wave that serves in computing the increment step.

We investigated a coupling approach between the level set method and LSM where the initialization is done using a crude estimate provided by the linear sampling method. The obtained results validate the efficiency of this coupling in the case of simply and multiply connected obstacles that are well separated.

### 6.3. Shape and topology optimization

#### 6.3.1. Incorporating manufacturing constraints in topology optimization

**Participant:** Grégoire Allaire.

With G. Michailidis and F. Jouve we study how to incorporate manufacturing constraints in topology optimization of structures using the level set method. The goal is to obtain a structure with optimal mechanical behaviour, which at the same time respects some predefined constraints imposed by the fabrication process. In this way, the final optimal shape is manufacturable and thus the method of shape and topology optimization turns to be industrially applicable.

The first constraints we have tackled are related to the limits of thickness a structure is forced to respect. We need to avoid optimal shape that contain very thin or thick members or even members that are very close between them. To achieve this, we have adopted two different approaches, a geometrical and a mechanical one. In the geometrical one, we have made extensive use of the notion of the signed-distance function to a domain. We have formulated a global constraint which guarantees that, at the end of the optimization process, the optimal structure respects the thickness limits. In the mechanical approach, we have tried to simulate the solidification process of a structure constructed via casting. We have set a time contraint, i.e. we have required that the structure cools earlier than some predefined time limit. We have started working on a more complicated thermal equation, a non-linear model with phase change, in order to describe more accurately the solidification process.

#### 6.3.2. Optimization of composite materials draping

**Participant:** Grégoire Allaire.

With G. Delgado we work on the optimization of composite materials draping. These composite structures are constructed by lamination of a sequence of unidirectional reinforced layers or plies. Each ply is typically a thin sheet of carbon fibers impregnated with polymer matrix material. The optimization variables are the geometries of these layers and they are parameterized by a level set function. In a first instance, we treat the problem of mass minimization (with a constraint on the maximal compliance) for a cantilever type composite structure, laminated with four layers of a given orthotropic material at different angles. The elasticity analysis is performed with the software Freefem++, coupled with a C++ routine to solve, by a finite difference scheme, the evolution of the level sets.

#### 6.3.3. A hybrid optimization method

**Participant:** Grégoire Allaire.
With Ch. Dapogny and P. Frey we develop a new method of geometric optimization for structures that relies on two alternative descriptions of shapes: on the one hand, they are exactly meshed so that mechanical evaluations by finite elements are accurate; on the other hand, we resort to a level-set characterization to describe their deformation along the shape gradient. The key ingredient is a meshing algorithm for building a mesh, suitable for numerical computations, out of a piecewise linear level-set function on an unstructured mesh. Therefore, our approach is at the same time a geometric optimization method (since shapes are exactly meshed) and a topology optimization method (since the topology of successive shapes can change thanks to the power of the level-set method). Our first results in 2-d have been announced. We continue to work on the 3-d case.

6.3.4. DeHomogenization

**Participant:** Olivier Pantz.

In most shape optimization problems, the optimal solution does not belong to the set of genuine shapes but is a composite structure. The homogenization method consists in relaxing the original problem thereby extending the set of admissible structures to composite shapes. From the numerical viewpoint, an important asset of the homogenization method with respect to traditional geometrical optimization is that the computed optimal shape is quite independent from the initial guess (even if only a partial relaxation is performed). Nevertheless, the optimal shape being a composite, a post-treatment is needed in order to produce an almost optimal non-composite (i.e. workable) shape. The classical approach consists in penalizing the intermediate densities of material, but the obtained result deeply depends on the underlying mesh used and the details level is not controllable. We proposed in [40] a new post-treatment method for the compliance minimization problem of an elastic structure. The main idea is to approximate the optimal composite shape with a locally periodic composite and to build a sequence of genuine shapes converging toward this composite structure. This method allows us to balance the level of details of the final shape and its optimality. Nevertheless, it was restricted to particular optimal shapes, depending on the topological structure of the lattice describing the arrangement of the holes of the composite. We lifted this restriction in order to extend our method to any optimal composite structure for the compliance minimization problem in [39]. Since, the method has been improved and a new article presenting the last results is in preparation. Moreover, we intend to extend this approach to other kinds of cost functions. A first attempt, based on a gradient method, has been made. Unfortunately, it was leading to local minima. Thus a new strategy has to be worked out. It will be mainly based on the same ideas than the one developed for the compliance minimization problem, but some difficulties are still to be overcome.

6.3.5. Level-Set Method

**Participant:** Olivier Pantz.

We have begun to work, with Gabriel Delgado, on a new level-set optimization method, based on a gradient method. The key idea consists in computing directly the derivative of the discretized cost functions. The main advantage is that it is usually more simple to implement than the standard approach (consisting in using a discretized version of the gradient of the cost function). Moreover, the results obtained are as good or even better than the one obtained in previous works. Nevertheless, this method has its drawbacks, since the cost function is only derivable almost everywhere (the zero level-set has to be transverse to the triangulation of the mesh). It follows that convergence toward the minimum by the gradient method is not granted. To overcome this problem, we intend to use a mix-formulation for the state function. An article is in preparation on this subject.

6.3.6. Robust Optimization

**Participant:** Olivier Pantz.

One of the main problem in shape optimization problems is due to the fact that the gradient is never computed exactly. When the current solution is far from a local optimum, this is not a problem: even a rough approximation of the gradient enable us to exhibit a descent direction. On the contrary, when close to a local optimal, a very precise computation of the gradient is needed. We intend, with G. Delgado, to use a-posteriori error estimates evaluate the errors made on the computation of the gradient and to ensure that at each step, a genuine descent direction is used in the gradient method.
6.3.7. **Optimization of a sodium fast reactor core**  
**Participants:** Grégoire Allaire, Olivier Pantz.

In collaboration with D. Schmidt, G. Allaire and E. Dombre, we apply the geometrical shape optimization method for the design of a SFR (Sodium Fast reactor) core in order to minimize a thermal counter-reaction known as the sodium void effect. In this kind of reactor, by increasing the temperature, the core may become liable to a strong increase of reactivity $\rho$, a key-parameter governing the chain-reaction at quasi-static states. We first use the 1 group energy diffusion model and give the generalization to the 2 groups energy equation. We then give some numerical results in the case of the 1 group energy equation. Note that the application of our method leads to some designs whose interfaces can be parametrized by very smooth curves which can stand very far from realistic designs. We don’t explain here the method that it would be possible to use for recovering an operational design but there exists several penalization methods that could be employed to this end. This work was partially sponsored by EDF. Our results will be published in the proceedings of the CEMRACS’11, during which part of the results have been obtained.

6.4. **Asymptotic models**

6.4.1. **Inverse scattering problem for coated obstacles**  
**Participants:** Nicolas Chaulet, Houssem Haddar.

In collaboration with L. Bourgeois, we considered the inverse scattering problem consisting in the identification of both an obstacle and its “equivalent impedance” from farfield measurements at a fixed frequency. The first specificity of this work is to consider the cases where this impedance is not a scalar function but a second order surface operator. The latter can be seen as a more general model for effective impedances and is for instance widely used for scattering from thin coatings. The second specificity of this work is to characterize the derivative of a least square cost functional with respect to this complex configuration. We provide in particular an extension of the notion of shape derivative to the cases where the impedance parameters cannot be considered as the traces of given functions. For instance, the obtained derivative does not vanish (in general) for tangential perturbations. The efficiency of considering this type of derivative is illustrated by some 2D numerical experiments based on a (classical) steepest descent method. The feasibility of retrieving both the obstacle and the impedance functionals is discussed in further numerical experiments [33].

6.4.2. **Interface conditions for thin dielectrics**  
**Participant:** Houssem Haddar.

Jointly with B. Delourme and P. Joly we established transmission conditions modelling thin interfaces that has (periodic) rapid variations along tangential coordinates. Motivated by non-destructive testing experiments, we considered the case of cylindrical geometries and time harmonic waves. We already obtained a full asymptotic description of the solution in terms of the thickness in the scalar case using so called matched asymptotic expansions. This asymptotic expansion is then used to derive generalized interface conditions and establish error estimates for obtained approximate models [15]. The analysis of the approximate problem for Maxwell’s equations is the subject of a forthcoming publication.

6.4.3. **Homogenization**  
**Participant:** Grégoire Allaire.

With I. Pankratova and A. Piatnitski we considered the homogenization of a non-stationary convection-diffusion equation posed in a bounded periodic heterogeneous domain with homogeneous Dirichlet boundary conditions. Assuming that the convection term is large, we give the asymptotic profile of the solution and determine its rate of decay. In particular, it allows us to characterize the “hot spot”, i.e., the precise asymptotic location of the solution maximum which lies close to the domain boundary and is also the point of concentration. Due to the competition between convection and diffusion, the position of the “hot spot” is not always intuitive as exemplified in some numerical tests.
With Z. Habibi we studied the homogenization of heat transfer in periodic porous media where the fluid part is made of long thin parallel cylinders, the diameter of which is of the same order than the period. The heat is transported by conduction in the solid part of the domain and by conduction, convection and radiative transfer in the fluid part (the cylinders). A non-local boundary condition models the radiative heat transfer on the cylinder walls. To obtain the homogenized problem we first use a formal two-scale asymptotic expansion method. The resulting effective model is a convection-diffusion equation posed in a homogeneous domain with homogenized coefficients evaluated by solving so-called cell problems where radiative transfer is taken into account. In a second step we rigorously justify the homogenization process by using the notion of two-scale convergence. One feature of this work is that it combines homogenization with a 3D to 2D asymptotic analysis since the radiative transfer in the limit cell problem is purely two-dimensional. Eventually, we provide some 3D numerical results in order to show the convergence and the advantages of our homogenization method.

6.4.4. Modelling and simulation for underground nuclear waste storage.

**Participants:** Grégoire Allaire, Harsha Hutridurga.

In the framework of the GDR MOMAS (Groupement de Recherches du CNRS sur les MOdélisations MATHématiques et Simulations numériques liées aux problèmes de gestion des déchets nucléaires) I am working with R. Brizzi, H. Hutridurga, A. Mikelic and A. Piatnitski on upscaling of microscopic models by homogenization (i.e. finding macroscopic models and effective coefficients).

We studied the Taylor dispersion of a contaminant in a porous medium. The originality of the model is that it takes into account surface diffusion and convection on the pores boundaries. We rigorously obtained the homogenized equation and studied the behavior of the effective dispersion tensor when varying various parameters.

In collaboration with a team of chemists (around J.-F. Dufrêche from the GNR Paris), we have undertaken the rigorous homogenization of a system of PDEs describing the transport of a N-component electrolyte in a dilute Newtonian solvent through a rigid porous medium. The motion is governed by a small static electric field and a small hydrodynamic force, which allowed us to use O’Brien’s linearized equations as the starting model. Convergence of the homogenization procedure was established and the homogenized equations were discussed. Based on the rigorous study of the underlying equations, it was proved that the effective tensor satisfies Onsager properties, namely is symmetric positive definite. This result justified the approach of many authors who used Onsager theory as a starting point.

6.4.5. A new membrane/plate modeling

**Participant:** Olivier Pantz.

Using a formal asymptotic expansion, we have proved with K. Trabelsi, that non-isotropic thin-structure could behave (when the thickness is small) like a shell combining both membrane and bending effects. It is the first time to our knowledge that such a model is derived. An article on this project is in preparation.

6.4.6. A new Liouville type Rigidity Theorem

**Participant:** Olivier Pantz.

We have recently developed a new Liouville type Rigidity Theorem. Considering a cylindrical shaped solid, we prove that if the local area of the cross sections is preserved together with the length of the fibers, then the deformation is a combination of a planar deformation and a rigid motion. The results currently obtained are limited to regular deformations and we are currently working with B. Merlet to extend them. Nevertheless, we mainly focus on the case where the conditions imposed to the local area of the cross sections and the length of the fibers are only “almost” fulfilled. This will enable us to derive rigorously new non linear shell models combining both membranar and flexural effects that we have obtained using a formal approach.

6.4.7. Lattices

**Participant:** Olivier Pantz.
With A. Raoult and N. Meunier (Université Paris Descartes), we have computed the asymptotic limit of a square lattice with three-points interactions. An article currently under review has been submitted on this work.

6.4.8. **Homogenization of axon Bundles**

**Participant:** Olivier Pantz.

With E. Mandonnet (Lariboisière Hospital), we have developed a new modeling for bundles of axons using homogenization technique. Previous works only focus (even if not explicitly) in the low density case: That is when the axon density is small. The aim is to determine which kind of electrical stimulation could trigger a signal into the axon. Under the low density assumption, the external electric field is independent of the membrane potential of the axon. If not, both are strongly coupled. Moreover, we have performed numerical simulations to determine what is the best position of the electrodes to enable the activation of the axons. This work has lead to the publications of an article [20] and a technical report [29]. Finally, we have begun to investigate more realistic modelings of the ionic flux based on the works of FitzHugh-Nagumo with a student, Xinxin Cheng, who spent three months at the CMAP.

6.5. **Diffusion MRI**

6.5.1. **Homogenized diffusion tensor and approximate analytical formulae for the long time ADC**

**Participants:** Jing-Rebecca Li, Houssem Haddar.

We model the bulk magnetization in biological tissue due to a diffusion gradient at the voxel level by a two compartment Bloch-Torrey partial differential equation. The cell membranes are modeled as infinitely thin permeable interfaces. We show the simulated long time apparent diffusion tensor of the PGSE sequence is close to the effective diffusion tensor from homogenization theory for both isotropic and anisotropic diffusion. For nearly isotropic diffusion we give analytical approximate formulae for the long time apparent diffusion coefficient in two and three dimensions. The approximate formulae allow us to robustly estimate the change in the cellular volume fraction from ADC measurements before and after cell swelling if the cells are approximately uniform in size. We can also use the formulae to estimate the average cell size.

6.5.2. **General ODE model of diffusion MRI signal attenuation**

**Participants:** Jing-Rebecca Li, Hang Tuan Nguyen.

We model the magnetization in biological tissue due to a diffusion gradient by a two compartment Bloch-Torrey partial differential equation with infinitely thin permeable membranes. We formulate a ODE model for the magnetization and show the simpler ODE model is a good approximation to the Bloch-Torrey PDE model for a variety of gradient shapes. Using the ODE model we determine of the change in the cellular volume fraction from the signal attenuation obtained before and after cell swelling. This method requires only the ADC and Kurtosis of the two signal attenuations and the numerical solution of an ODE system.
5. New Results

5.1. Model Design

5.1.1. Sensitivity Analysis of Complex Biophysical Models

5.1.1.1. Background

Sensitivity analysis (SA) is a fundamental tool in the building, use and understanding of mathematical models [44]. Sampling-based approaches to uncertainty and sensitivity analysis are both effective and widely used [38]. For this purpose, Sobol’s method is a key one [47]. Since it is based on variance decomposition, the different types of sensitivity indices that it estimates can fulfill different objectives of sensitivity analysis: factor prioritization, factor fixing, variance cutting or factor mapping [37]. It is a very informative method but potentially computationally expensive [38]. Besides the first-order effects, Sobol’s method also aims at determining the levels of interaction between parameters [48]. In [46], the authors also devised a strategy for sensitivity analysis that could work for correlated input factors, based on the first-order and total-order index from variance decomposition.

5.1.1.2. Algorithm numerical implementation

Computational methods to evaluate Sobol indices sensitivity rely on Monte-Carlo sampling and re-sampling [47], [40]. For k dimensional factor of model uncertainty, the k first-order effects and the ‘k’ total-order effects are rather expensive to estimate, needing a number of model evaluations strictly depending upon k [43]. Therefore, it is crucial to not only devise efficient computing techniques, in order to make best use of model evaluations [45], but also to have a good control of the estimation accuracy with respect to the number of samples.

With the objective of an efficient computational method for sensitivity analysis of functional-structural tree growth models, we proposed a new estimator based on Homma-Saltelli method to compute Sobol indices, which improves slightly their use of model evaluations thanks to a more balanced resampling strategy. This new estimator can be considered as an effort to improve the efficiency of SA methods for models. We also derived a theoretical analysis of the error estimation for the sensitivity analysis for the studied class of Sobol’s estimators (it can be applied to all the three Sobol’s estimators mentioned in this paper) with respect to the sampling size and the number of model evaluations. An analytical test function is used to test the error estimation, and we obtained that the error estimation in this paper gives out a better ‘upper bound’ than the previous works related to this problem. This error estimation directly relates to the variance of the result, so it can also be used for checking the confidence interval, which is usually difficult to attain.

We derived a theoretical analysis of the error estimation for the sensitivity analysis for the studied class of Sobol’s estimators (it can be applied to all the three Sobol’s estimators mentioned in this paper) with respect to the sampling size and the number of model evaluations. An analytical test function is used to test the error estimation, and we obtained that the error estimation in this paper gives out a better ‘upper bound’ than the previous works related to this problem. This error estimation directly relates to the variance of the result, so it can also be used for checking the confidence interval, which is usually difficult to attain.

The result has been accepted to be published in J of Rel. Eng. and Sys. Safety. Based on the published result for first order index of Sobol’s, we also extended this method to the second order index computing which is very important for us to know the precise pairs parameters with interactions between them, with the objective of making best use of the input-output model evaluation matrices that involve in the main part of the sensitivity analysis computing cost. Besides the computing efficiency improvement, one side result we got from the algorithm for second order index is that it can also make the final index has smaller variance so that the result can be more reliable.
Besides Sobol’s method, we also tried Morris method to complete the aim of ‘screening’ parameter of sensitivity analysis.

5.1.1.3. Strategies for FSPM

Global sensitivity analysis (SA) has an important role to play in functional-structural plant growth modeling by assessing the different source of uncertainty help us to gain some insights inner the models so as to explain the behavior of them. Different FSPMs have different scales of model design, which leads to all types of diverse multi-biophysical processes.

To study specifically how global SA can help for FSPMs, SA was applied on a wide variety of functional-structural plant models, typically the 3 FSPMs: firstly a simple source-sink model of maize growth, is used to specifically study the process of carbon (C) allocation among expanding organs during plant growth, with simple plant structure, multi-stage and detailed observations, secondly the GreenLab model of tree growth (applied to poplar tree) characterized by the retroaction of plant functioning on its organogenesis [41], which describes tree structural plasticity in response to trophic competition, lastly a functional-structural model, NEMA [16], describing C and nitrogen (N) acquisition by a wheat plant as well as C and N distributions between plant organs after flowering. This model has the specificity to integrate physiological processes governing N economy within plants: root N uptake is modeled following the transport systems high affinity transport systems (HATS) and low affinity transport systems (LATS), and N is distributed between plant organs according to the turnover of the proteins associated to the photosynthetic apparatus. C assimilation is predicted from the N content of each photosynthetic organ. Consequently, this model is more mechanistic but also more complex than the two previous ones. Another objective is to explore an effective simulation design to help the sensitivity analysis for complicate models with several logically distinct but biological functioning interacted moduls.

All these SA result shadowed a light to the models for us to diagnosis the model behavior and will bring a big step for parameter estimation and experimental simplification in our modelling next.

5.1.2. A model for Cecropia sciadophylla under fluctuating environmental conditions

In collaboration with Patrick Heuret (INRA, JRU Ecofog, Kourou, French Guiana), we developed a tree growth model dedicated to *Cecropia sciadophylla*, a neotropical species from the genus Cecropia. These trees have interesting properties from a modeller’s point of view: they have a simple architecture, their number of phytomers remain limited even for old individuals, and most importantly, [39] and [49] have developed a methodology based on morphological observations to estimate tree age on *C. obtusa* and *C. sciadophylla* respectively. It is therefore possible to fully describe the tree structure and topology from morphological observations, which is very uncommon for trees: for most tree species, their high stature, complex structure, and long life span drastically increase the fieldwork required to collect data at the organ scale and hamper the development, calibration and validation of functional-structural tree growth models and their potential applications in the field of forest management.

We used datasets collected on 18 trees in 2007 and 2008 in French Guiana to develop and evaluate our model. Our objective was to analyse the influence of fluctuating environmental conditions on the dynamics of trophic competition within *C. sciadophylla* trees. We defined an integrated environmental factor that includes meteorological medium-frequency variations and a relative index representing the local site conditions for each plant. The meteorological variations were input from pluviometry data, that could be considered as the main fluctuating environmental stress under that tropical climate. The relative index was estimated based on inversion of our model using data from respectively 11 trees for model calibration (those measured in 2007) and 7 trees for model evaluation (those measured in 2008). This study provided a model that can be seen as a tool to disentangle the ontogenic variations (low-frequency trend) and the environmental variations (medium-frequency variation). One paper was accepted for publication [22].
5.1.3. Using model inversion to analyze the effects of inter-tree competition on four Pine trees grown under two contrasted density conditions

In collaboration with Guo Hong and Lei Xiangdong (Chinese Academy of Forestry, Beijing, China), we analyzed the characteristics of individual tree response to competition on source-sink balance through the calibration of the GreenLab model.

Four Chinese pine trees (Pinus tabulaeformis Carr.) were destructively measured in November 2009 from the nursery garden located in the Yuanyiqi forest farm, Beijing, China. Two 13-year-old trees (T1 and T2) were from a high density plantation (3500N/ha) and two 10-year-old trees (T3 and T4) were from a low density plantation (2000N/ha). We first examined the statistical differences in the tree morphologies and topologies. Significant differences were found for internode diameter, internode biomass and needle biomass between the two densities, but not for internode length. In a second step, we studied the ability of the GreenLab model to simulate the plasticity of pine trees grown under different densities. To fulfil these objectives, it was necessary to find a way to characterize the competition conditions of each tree. Given the inherent difficulty of identifying the most relevant experimental measurements for this characterization, we proposed to represent the effects of competition on the tree growth through a single tree-specific parameter of GreenLab, called characteristic surface area, and to estimate it for each tree by model inversion, together with the more classical endogenous species parameters. This will eventually allow us to examine whether the obtained value of this characteristic surface area could be correlated to other possible indicators of competition pressure. This could pave the way to the development of an individual-based stand growth model including the effects of a competition index.

One paper was submitted to Trees - Structure and Functions.

5.1.4. Coffee trees and genetics

In collaboration with Sylvie Sabatier (INRA, AMAP), Philippe de Reffye (CIRAD, AMAP) and Perla Hamon (IRD Montpellier), we studied the architectural and genetic diversities in 5 Coffea species, native from Madagascar. We explore two complementary methods: the genetic diversity using molecular markers (genomic- and/or EST-microsatellites) and the variability of adaptive traits between populations with different ecological niches. We focused on 5 Coffea species endemic to Madagascar, some of which are classified as critically endangered in the World Conservation Union(IUCN) Red List. For each species, architecture and genetic comparative analyses between individuals growing in situ (natural forest) and ex situ (common garden test) are being performed. In parallel, the same populations are analysed using the GreenLab model. These results will be used to study the potential links between the parameters of GreenLab and the allelic distribution in these populations. This is the subject of the PhD of Domohina Andrianasolo (CIRAD, Montpelier and FOFIGA, Antananarivo, Madagascar). This work was presented at the XVIII International Botanical Congress (IBC) [34].

5.1.5. Methods for tree crown analysis and application to young Eucalyptus

Based on the pioneer work on coffee trees of Philippe de Reffye, a stochastic model was developed to describe the topological development of trees. In the model, growth and branching processes are driven by the respective probabilities of activity, rest or death of apical and lateral buds. Because of its mathematical formulation, the model inversion can be done analytically which is rare and parameter values can be estimated from experimental data. The MATLAB software GLOUPS developed by Philippe de Reffye was used. We explored the feasibility of calibrating this stochastic model for eucalyptus, which presents the additional difficulty of a continuous growth with no marked endogenous cessation. Incomplete systems were also defined for the case, common with trees, of incomplete datasets. An adequate strategy was defined to sample measurements and applied to five eucalyptus trees (data collected by Pr Lei Xiangdong, Guo Hong and Diao Jun, Chinese Academy of Forestry, Beijing, China).

One paper was accepted for publication [20].

5.2. Model Evaluation and Parameter Estimation
5.2.1. Maximum Likelihood Estimation

In [18], a first approach for parameter estimation was introduced based on the assumption of an underlying deterministic model of biomass production and uncorrelated errors in the mass measurements of different types of organs in the plant structure. A novel idea is developed on the modeling plant growth in the framework of non-homogeneous hidden Markov models, for a certain class of plants with known organogenesis (structural development). Unknown parameters of the models are estimated via a stochastic variant of a generalised EM (Expectation-Maximization) algorithm where both steps (E-step, M-step) are non-explicit. For this reason, the E-step is approximated via a sequential Monte-Carlo procedure (sequential importance sampling with resampling) and the M-step is separated into two steps (Conditional-Maximization), where before applying a numerical maximization procedure (quasi-Newton type), a large subset of unknown parameters is updated explicitly conditioned on the other subset. The model is tested with real data and the results are satisfying. Further work is in progress, including MCMC techniques for parameter estimation (with the collaboration of Dr. Sonia Malefaki from the University of Patras, Greece) and Bayesian type estimation, see [33].

5.2.2. Convolution Particle Filter for parameter estimation

Although Kalman filter is applied to various fields and dominated for decades, it is limited by its assumptions of linearity. Particle filter, which combines Bayesian inference with Monte Carlo sequential sampling approach, is a method using different combinations of random variables sampled directly from the parameter space (or the state space) to estimate parameters and states of a complex system. These combinations, generally called particles, propagate by introducing new observations and provide updated posterior distributions by taking into account their weights. Meanwhile, a resampling procedure is used to prevent the degeneracy problem. Since classical filtering methods are generally not able to estimate the dynamical state vector along with the unknown parameters, the convolution particle filter is implemented based on convolution kernel approximation to meet the need while modelling with Markovian dynamical system. Several tests are carried out to examine the performance of the Convolution Particle Filtering method [42], [36], and efforts are made to find the optimal perturbation parameters. The applications of the method rely on the Lotka-Volterra model and the sugar beet model. Since the quality of the estimations is limited by the number of the observations, the Conditional Iterative Bayesian Filtering method is applied. The principal is simply to use the posteriori distributions as the a priori information to re-perform over and over again the estimation algorithm and each time we introduce the same sequence of observations. This approach helps us to improve significantly the final estimation of the hidden states and the unknown parameters while testing with the virtual data. The bootstrapping method is implemented in order to compare with the results from different methods. In the case of applications based on dynamical stochastic systems, two types of noise are introduced, one is involved in the modelization technique and the other is attached to the observation procedure. An alternation of deterministic parameter estimation and stochastic parameter estimation is proposed (in progress) which allows us to estimate these two kinds of parameters at the same time.

5.2.3. Modelling the inter-individual variability of organogenesis in sugar beet populations

Modelling the inter-individual variability in plant populations is a very important issue to enhance the predictive capacity of plant growth models at the field scale. In the case of sugar beet, this variability is well illustrated by the phyllochron (the thermal time elapsing between the appearance of two successive leaves): if the mean phyllochron remains very stable across seasons, there is a high heterogeneity among individuals. Likewise, seedling emergence may strongly vary within a population, potentially inducing important variations in individual plant productions. A hierarchical segmented model was used to describe and study the variability of the dynamics of leaf appearance in sugar beet crops. The use of this nonlinear mixed model allows for a better handling of the heterogeneity in the plant population, and gives estimates of this variability: each model’s parameter is considered as a random variable, varying from one plant to another around a mean population value, with a given variance.
These mean population values and inter-individual variability can be used as input of functional-structural models, the main issue being then to compute the propagation of these sources of probabilistic uncertainty in the dynamic system of Greenlab.

5.3. Methods for the Applications

5.3.1. Optimization of Phenotyping based on a Parameter Selection Methodology

The model Cornflo is a functional plant growth model simulating Corn’s growth and yield. Based on it, the classification of environmental scenarios is researched in term of their influences to corn’s yield, and their parameter estimation capabilities for Cornflo parameters. The initial qualitative analysis of parameter estimation results shows that environmental scenarios’ classification benefit estimation accuracy and identifiability. Currently this project is researched from three aspects. Firstly, different clustering techniques are tested to find the most proper scenarios categories for parameter estimation. Secondly, the scenarios clusters are used for botanical experiments optimization, such as the selection of experimental locations. Lastly, parameter estimation is optimized and researched in a practical use for plant growth models.

5.3.2. Plant-Soil interaction and Optimal Control of Irrigation

This work is performed in collaboration with JC Mailhol (Cemagref). Irrigation scheduling is an important issue for crop management, in a general context of limited water resources and increasing concern about agricultural productivity. Methods to optimize crop irrigation should take into account the impact of water stress on plant growth and the water balance in the plant-soil-atmosphere system. For this purpose, different plant-soil interaction models are proposed to simulate the functional plant growth. In particular, a compartment plant model is designed to integrate water stress impact on different main physiological processes of crop: biomass production, biomass allocation, and foliar senescence. This model is applied and calibrated for maize, in order to predict the harvest index according to the stress undergone by crop during its whole cycle. As for the optimal control problem of irrigation, it can be formulated by considering a price for the crop yield and for the water resource. Dynamic programming is then applied to the plant-soil system to determine an optimal irrigation strategy.
6. New Results

6.1. Algorithmic study of linear functional systems

Participants: Alban Quadrat [correspondent], Thomas Cluzeau [ENSIL, Univ. Limoges].

In [77], [76], [97], it is shown that every linear functional system (e.g., PD systems, differential time-delay systems, difference systems) is equivalent to a linear functional system defined by an upper block-triangular matrix of functional operators: each diagonal block is respectively formed by a generating set of the elements of the system satisfying a purely \( i \)-codimensional system. Hence, the system can be integrated in cascade by successively solving (inhomogeneous) \( i \)-codimensional linear functional systems to get a Monge parametrization of its solution space [129] [89]. The results are based on an explicit construction of the grade/purity filtration of the module associated with the linear functional system. This new approach does not use involved spectral sequence arguments as is done in the literature of modern algebra [102], [103]. To our knowledge, the algorithm obtained in [77], [97] is the most efficient algorithm existing in the literature of non-commutative algebra. It was implemented in the PURITY FILTRATION package developed in Maple (see Section 5.7) and in the homalg package of GAP 4 (see Section 5.8).

Given a linear multidimensional system (e.g., ordinary/partial differential systems, differential time-delay systems, difference systems), Serre’s reduction aims at finding an equivalent linear multidimensional system which contains fewer equations and fewer unknowns. Finding Serre’s reduction of a linear multidimensional system can generally simplify the study of structural properties and of different numerical analysis issues, and it can sometimes help solving the linear multidimensional system in closed form. In [17], Serre’s reduction problem is studied for underdetermined linear systems of partial differential equations with either polynomial, formal power series or analytic coefficients and with holonomic adjoints in the sense of algebraic analysis [102], [103]. These linear partial differential systems are proved to be equivalent to a linear partial differential equation. In particular, an analytic linear ordinary differential system with at least one input is equivalent to a single ordinary differential equation. In the case of polynomial coefficients, we give an algorithm which computes the corresponding linear partial differential equation.

In [45], we give a complete constructive form of the classical Fitting’s lemma in module theory which studies the relation between equivalences of linear systems and isomorphisms of their associated finitely presented modules. The corresponding algorithms were implemented in the OREMORPHISMS package (see Section 5.5).

6.2. A new approach to classes of quasilinear PD systems

Participants: Alban Quadrat [correspondent], Thomas Cluzeau [ENSIL, Univ. Limoges], Daniel Robertz [Univ. Aachen].

Many partial differential systems appearing in mathematical physics, engineering sciences and mathematical biology are nonlinear. Unfortunately, algebraic analysis and \( D \)-module theory, which were successful for the algorithmic study of linear partial differential systems, cannot consider nonlinear PD systems. This project aims at developing a generalization of the algebraic analysis approach to certain classes of quasilinear partial differential systems appearing in mathematical physics and engineering sciences (e.g., Burgers’ equation, shallow water, Euler equations for an incompressible fluid, traffic flow, gas flow). In [44], we have shown how constructive methods of differential algebra and algebraic analysis could be combined to extend results obtained in [45] and [104] for linear PD systems and how they allowed us to compute conservation laws, internal symmetries and decompositions of the solution space of certain classes of quasilinear PD systems. The algorithms have been implemented in the JANETMORPHISMS package dedicated to this new approach (see Section 5.6).
6.3. Stabilization problems & Noncommutative geometry

**Participant:** Alban Quadrat [correspondent].

In [124], [123], [122], it was shown how the fractional representation approach to analysis and synthesis problems developed by Vidyasagar, Desoer, Callier, Francis, Zames..., could be recast into a modern algebraic analysis approach based on module theory (e.g., fractional ideals, algebraic lattices) and the theory of Banach algebras. This new approach successfully solved open questions in the literature. Basing ourselves on this new approach, we explain in [126] why the non-commutative geometry developed by Alain Connes is a natural framework for the study of stabilizing problems of infinite-dimensional systems. Using the 1-dimensional quantized calculus developed in non-commutative geometry and results obtained in [124], [123], [122], we show that every stabilizable system and their stabilizing controllers naturally admit geometric structures such as connections, curvatures, Chern classes... These results are the first steps toward the use of the natural geometry of the stabilizable systems and their stabilizing controllers in the study of the important $H_{\infty}$ and $H_2$-problems.

6.4. Stabilization of time-delay systems

**Participants:** Alban Quadrat [correspondent], Arnaud Quadrat [SAGEM, MASSY].

In [127], we study the stabilization problem of a linear system formed by a simple integrator and a time-delay. We show that the stabilizing controllers of such a system can be be rewritten as the closed-loop system defined by the stabilizing controllers of the simple integrator and a distributed delay. This result is used to study tracking problems appearing in the study of inertially stabilized platforms for optical imaging systems.

6.5. Singular boundary problems

**Participants:** Georg Regensburger [correspondent], Anja Korporal [RICAM, Linz], Bruno Buchberger [RISC, Linz], Markus Rosenkranz [Univ. Kent].

In [55], we present results on algorithmic methods for singular boundary problems for linear ordinary differential equations. Moreover, the implementation of integro-differential operators and the corresponding algorithms for boundary problems in the computer algebra system Maple is discussed. The operations implemented for regular boundary problems include computing Green’s operators as well as composing and factoring boundary problems. For singular boundary problems, compatibility conditions and generalized Green’s operators can be computed. In [94], we give a survey and new results on our algebraic and symbolic approach to boundary problem developed over the last years. The construction of integro-differential operators and polynomials over an integro-differential algebra is described in detail along with a generic implementation of the corresponding canonical forms and algorithms.

6.6. Model of reaction networks

**Participants:** Georg Regensburger [correspondent], Stefan Müller [RICAM, Linz].

In a joint work, Stefan Müller and G. Regensburger propose a notion of generalized mass action systems that could serve as a more realistic model for reaction networks in intracellular environments; classical mass action systems capture chemical reaction networks in homogeneous and dilute solutions, see e.g. [111] and [114]. We show that several results of Chemical Reaction Network Theory carry over to the case of generalized mass action kinetics. Our main result essentially states that, if the sign vectors of the stoichiometric and the kinetic-order subspace coincide, there exists a unique positive complex balancing equilibrium in every stoichiometric compatibility class.

6.7. Ruin probabilities

**Participants:** Georg Regensburger [correspondent], Hansjörg Albrecher [Univ. Lausanne], Corina Constantinescu [Univ. Lausanne], Zbigniew Palmowski [Univ. Wroclaw], Markus Rosenkranz [Univ. Kent].
In a cooperation with Hansjörg Albrecher, Corina Constantinescu (both University of Lausanne), Zbigniew Palmowski (University of Wroclaw), and Markus Rosenkranz (University of Kent), we developed a symbolic technique to obtain asymptotic expressions for ruin probabilities and discounted penalty functions in renewal insurance risk models when the premium income depends on the present surplus of the insurance portfolio. The analysis is based on boundary problems for linear ordinary differential equations with variable coefficients and the corresponding Green’s operators (integral operators), generalizing the approach from [99]. We obtain also closed-form solutions for more specific functions in the compound Poisson risk model.

6.8. Systems of linear ordinary integro-differential equations

Participants: Alban Quadrat, Georg Regensburger [correspondent].

A. Quadrat and G. Regensburger (in the frame of his grant) are working on a new approach for studying algebraic and algorithmic properties of systems of linear ordinary integro-differential equations with boundary conditions. In a recent series of papers, in particular, [100], [101], V. V. Bavula obtained numerous algebraic results for modules over the ring of integro-differential operators with polynomial coefficients using generalized Weyl algebras. We are interested in how far some of his approach can be made algorithmic and generalized to boundary problems. First results in this direction were presented at the Journées Nationales de Calcul Formel (JNCF 2011).

6.9. Stabilization of MIMO fractional systems with delays

Participants: Catherine Bonnet, Le Ha Vy Nguyen, Alban Quadrat.

In order to yield the set of all stabilizing controllers of a large class of MIMO fractional time-delay systems, we may look for coprime factorizations of the transfer function and their corresponding Bézout factors. As primary results, in considering $H_\infty$ stability, left coprime factorizations and left Bézout factors have been determined analytically from the transfer function. Then a particular stabilizing controller has been derived. We also proved the existence of right coprime factorizations.

6.10. Stability analysis of (fractional) delay systems of neutral type

Participants: Catherine Bonnet, André Fioravanti [UNICAMP], Le Ha Vy Nguyen.

The $H_\infty$-stability analysis of (fractional) delay systems of neutral type has been studied in [14]. It was shown that the chains of poles are asymptotic to vertical axes which position depend on the roots of a certain polynomial (easily determined from the given transfer function). In [14] the case of roots of multiplicity one was completely treated whereas the case of roots of multiplicity greater than one was considered only in the particular case of neutral systems whose transfer function is a product of transfer functions of neutral systems with one delay. This year, we have studied more deeply the case of roots of multiplicity greater than one for both, standard and fractional delay systems.

6.11. Matrix Norm Approach for Control of Linear Time-Delay Systems

Participants: Catherine Bonnet, André Fioravanti [UNICAMP], José Claudio Geromel [UNICAMP], Silviu Niculescu.

In [46], we have treated the time-delay linear systems control design in the framework of complete and partial information. We were able to find linear controllers that increase the first stability window imposing at the same time that the delay-free system is stable using some properties about the norms of the state-space matrices. Our method treated the design problem by numeric routines based on Linear Matrix Inequalities (LMI) arisen from classical linear time invariant system theory coupled together with a unidimensional search. Both the state and output feedback design, were solved.

6.12. Interval Observers

Participants: Frédéric Mazenc, Silviu Niculescu, Olivier Bernard [UNICAMP].
The technique, based on the notion of interval observer, is a recent state estimation technique, which offers the advantage of providing information on the current state of a system at any instant of time. The first interval observers were relying on the assumption that the system was cooperative or, roughly speaking, “almost” cooperative.

In the contribution [25], we have proved that, for any time-invariant exponentially stable linear system with additive disturbances, time-varying exponentially stable interval observers can be constructed. The technique of construction relies on the Jordan canonical form that any real matrix admits and on time-varying changes of coordinates for elementary Jordan blocks which lead to cooperative linear systems. We applied our to the case of linear systems with input and output that are detectable.

The paper [31] focused on the analysis and the design of families of interval observers for linear systems with a point-wise delay. First, we proved that classical interval observers for systems without delays are not robust with respect to the presence of delays, no matter how small delays are. Next, we have shown that, in general, for linear systems with delay, the classical interval observers endowed with a point-wise delay are unstable. A new type of design of interval observers enabling to circumvent these obstacles is proposed. It provides with framers that incorporate distributed delay terms. The proposed interval observers are assessed through a non-linear biotechnological model.

6.13. Partial Differential Equations
Participants: Frédéric Mazenc, Christophe Prieur [GIPSA Lab, CNRS].

In [33] and [65], for families of partial differential equations (PDEs) with particular stabilizing boundary conditions, we have constructed strict Lyapunov functions. The PDEs under consideration were parabolic and, in addition to the diffusion term, might contain a nonlinear source term plus a convection term. The boundary conditions were the classical Dirichlet conditions, or the Neumann boundary conditions or a periodic one. The constructions relied on the knowledge of weak Lyapunov functions for the nonlinear source term. The strict Lyapunov functions were used to prove asymptotic stability in the framework of an appropriate topology. Moreover, when an uncertainty is considered, our construction of a strict Lyapunov function made it possible to establish some robustness properties of Input-to-State Stability (ISS) type.

In [64], we have considered a family of time-varying hyperbolic systems of balance laws. The partial differential equations of this family can be stabilized by selecting suitable boundary conditions. For the stabilized systems, the classical technique of construction of Lyapunov functions provides with a function which is a weak Lyapunov function in some cases, but is not in others. We transform this function through a strictification approach which gives a time-varying strict Lyapunov function which allows us to establish asymptotic stability in the general case and a robustness property with respect to additive disturbances of Input-to-State Stability (ISS) type.

Participants: Frédéric Mazenc, Siviu Niculescu.

In the work [32], we propose a new approach for the stabilization of nonlinear time-varying forward-complete systems with delay in the input. This approach is a new reduction model approach, which relies on operators of a new type. It presents three advantages. First, the corresponding control laws do not include distributed terms. Second, it yields closed-loop systems with positive solutions that can be easily derived. Finally, the stabilized systems possess some robustness properties (for instance of ISS type) that can be estimated.

6.15. Backstepping with delay
Participants: Frédéric Mazenc, Siviu Niculescu, Mounir Bekaïk.
In the contributions [30] and [63], we revisited the problem of constructing globally asymptotically stabilizing control laws for nonlinear systems in feedback form with a known pointwise delay in the input. The result we obtained covers a family of systems wider than those studied in the literature and endows with control laws with a single delay, in contrast to those given in previous works which include two distinct pointwise delays or distributed delays. The strategy of design is based on the construction of an appropriate Lyapunov-Krasovskii functional. An illustrative example ends the paper.

6.16. Certifying good performance for adaptive systems

Participants: Frédéric Mazenc, Michael Malisoff [Louisiana State University].

The usual adaptive control problem is to design a controller that forces all trajectories of the system to track a prescribed trajectory, while keeping the estimator of the unknown constant parameter vector bounded. We studied the important and more difficult adaptive tracking and estimation problem of simultaneously (1) forcing the trajectories of the system to track a given trajectory and (2) identifying the parameter vector. This problem was known to be solvable when the regressor satisfies a persistency of excitation condition, but the known results did not provide a strict Lyapunov function for the augmented error dynamics and so could not certify good performance such as ISS with respect to uncertainty added to the controller.

Our main result from [118] covers adaptively controlled first-order nonlinear systems that satisfy the persistency of excitation condition and are affine in the unknown parameter vector. Our contribution consists in particular in constructing global strict Lyapunov functions for the augmented error dynamics. In [28], we extended [118] to adaptive tracking for nonlinear systems in feedback form with multiple inputs and unknown high-frequency gains multiplying the controllers. The control gains must be identified as part of the control objective. High-frequency gains are important for electric motors, flight dynamics, and robot manipulators. We used a persistency of excitation condition that again ensured tracking and parameter identification and led to the explicit construction of a global strict Lyapunov function for the closed loop augmented error dynamics. The strict Lyapunov function was key to proving integral ISS with respect to time varying uncertainties added to the unknown parameters.

6.17. Improved Model Predictive Control design

Participants: Sorin Olaru, Morten Hovd [NTNU, Trondheim, Norway].

New results [38] have been obtained toward the computation of feasible sets for linear model predictive control techniques, based on set relations and not on the conventional orthogonal projection. Further, the problem of computing suitable inner approximations of the feasible sets was considered. Such approximations are characterized by simpler polytopic representations, and preserve essential properties as convexity, positive invariance, inclusion of the set of expected initial states.

6.18. Particle Swarm Optimization for reduced order Hinfini controllers with constraints handling

Participants: Guillaume Sandou, Gilles Duc [Supelec (E3S) Control Department], Mohamed Yagoubi [Ecole des Mines de Nante].

Efficient dedicated methods have been developed for Hinfinity controller synthesis. However, such methods require translating the design objectives using weighting filters, whose tuning is not easy; in addition they lead to high order controllers which have to be reduced. A particle swarm optimization method is used to solve both problems successively: after having optimized the filters according to the design specifications using a full order controller but no reformulation of the constraints, a reduced-order one is computed using the obtained filters. Experimental tests for a pendulum in the cart exhibit much than satisfactory results [95], [52]. In addition, the design of Hinfinity static output feedback has been done and tested using the Compleib library and exhibiting similar results to those obtained with the HIFOO solver [86].
6.19. Ant colony for symbolic regression

**Participants:** Guillaume Sandou, Bianca Minodora Heiman.

The identification of systems is a key feature to get representative models and so to design efficient control laws. Numerous methods exist to identify the parameters of nonlinear systems when the global structure of the model is given. The problem appears to be much more difficult when this structure is unknown (symbolic regression). We introduce ant colony optimization (ACO) in solving the problem of nonlinear systems identification in the case of an unknown structure of the model [53]. Numerical results prove the viability of the approach.

6.20. Robust optimization for energy management

**Participants:** Guillaume Sandou, Henri Borsenberger, Philippe Dessante [Supelec (E3S) Energy Department].

Many studies have considered the solution of Unit Commitment problems for the management of energy networks. In this field, earlier work addressed the problem in deterministic cases and in cases dealing with demand uncertainties. In this paper [15], the authors develop a method to deal with uncertainties related to the cost function. Indeed, such uncertainties often occur in energy networks (waste incinerator with a priori unknown waste amounts, cogeneration plant with uncertainty of the sold electricity price...). The corresponding optimization problems are large-scale stochastic non-linear mixed integer problems. The developed solution method is a recourse-based programming one. The main idea is to consider that amounts of energy to produce can be slightly adapted in real time, whereas the on/off statuses of units have to be decided very early in the management procedure. Results show that the proposed approach remains compatible with existing Unit Commitment programming methods and presents an obvious interest with reasonable computing loads.

6.21. Modeling and control of Acute Myeloid Leukemia

**Participants:** José Luis Avila Alonso, Annabelle Ballesta [BANG project-team], Frédéric Bonnans [COMMANDS project-team], Catherine Bonnet, Jean Clairambault [BANG project-team], Xavier Dupuis [COMMANDS project-team], Pierre Hirsch [INSERM Paris (Team 18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Jean-Pierre Marie [INSERM Paris (Team 18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Faten Merhi [INSERM Paris (Team 18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Silviu Niculescu, Hitay Özbay [Bilkent University, Ankara, Turkey], Ruoping Tang [INSERM Paris (Team 18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris].

In order to better take into account physiological phenomena as well as better understand the effect of the new anti-FLT3 therapy for AML, we have modified the model M. Adimy and F. Crauste in two ways [88]:
- we have introduced a modeling of quick self-renewal of cells in each stage of maturation.
- we have modeled each phase of the proliferating compartment (that is $G_1$, $S$, $G_2$, and $M$) separately. For the time being, only the $M$-phase is supposed to have a fixed time duration as it is well-known that the short time necessary to perform mitosis is hardly submitted to any variation.

In parallel to this modeling task, Faten Merhi and Annabelle Ballesta have performed experiments in order to identify parameters of the model. These experiments (which will continue in 2012) tend to show that we will converge to a patient-dependant model.

6.22. Chronic myelogenous leukemia

**Participants:** Frédéric Mazenc, Silviu Niculescu, Peter Kim [Univ. of Sydney].
The paper [26] focuses on the stability analysis of a delay-differential system encountered in modeling immune dynamics during imatinib treatment for chronic myelogenous leukemia (CML). A simple algorithm is proposed for the analysis of delay effects on the stability. Such an algorithm takes advantage of the particular structure of the dynamical interconnections of the model. The analysis shows that the model yields three fixed points, two of which are always unstable and one of which is sometimes stable. The stable fixed point corresponds to an equilibrium solution in which the leukemia population is kept below the cytogenetic remission level. This result implies that, during imatinib treatment, the resulting anti-leukemia immune response can serve to control the leukemia population. However, the rate of approach to the stable fixed point is very slow, indicating that the immune response is largely ineffective at driving the leukemia population towards the stable fixed point. To extend the stability analysis with respect to the delay parameter, we conduct a global nonlinear analysis to demonstrate the existence of unbounded solutions. We provide sufficient conditions based on initial cell concentrations that guarantee unbounded solutions and comment on how these conditions can serve to predict whether imatinib treatment will result in a sustained remission based on a patient’s initial leukemia load and initial anti-leukemia T cell concentration.

6.23. Control of a model of human heart

Participants: Frédéric Mazenc, Michael Malisoff [Louisiana State University], Marcio de Queiroz [Louisiana State University].

We worked on the control of human heart rate during exercise [27], which is a problem that has implications for the development of protocols for athletics, assessing physical fitness, weight management, and the prevention of heart failure. We provided new stabilization techniques, based on the notion of tracking, for a recently-proposed nonlinear model for human heart rate response that describes the central and peripheral local responses during and after treadmill exercise. The control we proposed respect the sign constraint imposed by the model and we proposed observers to cope with the case (important from a practical point of view) where some of the variables are not measured.

6.24. Marine Robotic Surveys

Participants: Frédéric Mazenc, Michael Malisoff [Louisiana State University], Fumin Zhang [Georgia Tech.].

The works [24], [61] was inspired by the recent Deepwater Horizon oil spill disaster. The goal was to develop and implement robotic surveying methods to evaluate the immediate and longer term environmental impacts of the oil spills. It was joint with Michael Malisoff from the LSU and a Georgia Tech robotics team led by Fumin Zhang. Robotic surveying methods provide a low cost and convenient way to collect data in marsh areas that are difficult to access by human based methods. We designed strict Lyapunov functions that made it possible to use ISS to quantify the robustness of collision avoiding curve tracking controllers under controller uncertainty. The controllers are designed to keep the robot a fixed distance from, but moving parallel to, a two dimensional curve. Four challenges in applying ISS to curve tracking are (a) the need to restrict the magnitudes of the uncertainty to keep the state in the state space and build a strict Lyapunov function, (b) the likelihood of time delays in the controllers in real time applications, (c) possible parameter uncertainty such as unknown control gains, and (d) generalizations to three dimensional curve tracking. We overcame challenge (a) by finding maximum bounds on the perturbations that maintain forward invariance of a nested family of hexagons that fill the state space and transforming a nonstrict Lyapunov function into a strict Lyapunov function on the full state space. To address challenge (b), we used a Lyapunov-Krasovskii approach from [119] to convert the strict Lyapunov function into a Lyapunov-Krasovskii functional. This led to an upper bound on the admissible controller delay that can be introduced into the controller while still maintaining ISS.
Galen Team

6. New Results

6.1. Reconstruction

Participants: Panagiotis Koutsourakis, Helene Langet, Loic Simon, Olivier Teboul, Gilles Fleury, Elisabeth Lahalle, Yves Trousset, Cyril Riddell, Nikos Paragios.

- **Image-based Procedural Modeling of Urban environments**: In [20] we develop a multiple hypotheses testing algorithm for image-based/grammar-driven building modeling. Shape grammars are used to express the variation of the observed architecture. Such a model is coupled with the observations through a maximum likelihood principle where the aim is to maximize the posterior segmentation probability in the image plane given the partition being determined from the grammar derivation. The unknown parameters of the process involve the grammar derivation tree and the associated parameters. Such a mixed continuous/discrete problem is solved through a hill climbing approach that involves joint perturbations in the derivation and parameter space. Promising results demonstrated the potentials of such a formulation for complex Parisian architectures. This idea was further extended in [40] where reinforcement learning was used as optimization principle. Performance in particular computational gain over [20] demonstrated the extreme potentials of such a formulation. In order to cope with multi-view geometry, the grammar was further derived to include 3D components and the optimization process was amended to deal with multiple views. An evolutionary computation process (based on consistent mutation and recombination of partial grammar trees) was proposed to fuse image and depth-based information. The use of the Pareto frontier between the two concurrent components of the objective function provides a principle way to determine the optimal solution of the designed objective function.

- **Compressed Sensing Digital Subtraction Rotational Angiography**: in [39] we develop an extension of iterative filtered backprojection method for reconstruction of three-dimensional vascular structures from two spins. Our contribution refers to an approach that improves the reconstruction quality of non-sparse volumes when there exists a sparse combination of these volumes. This is achieved through a joint reconstruction of the mask and contrast volumes via $l^1$-minimization of sparse priors. These ideas were further explored to address three-dimensional reconstruction in interventional radiology in [30] through a regularized extension of the iterative filtered backprojection algorithm. To this end the conventional TV-norm was replaced from a new sparsity constraint that relies on the $l^1$-minimization-norm and the positivity constraint. The use of such a constraint allows for removing most of the subsampling artifacts while preserving background structures.

6.2. Matching/Segmentation

Participants: Haimhem Boussaid, Iasonas Kokkinos, Chaohui Wang, Bo Xiang, Ahmet Besbes, Ben Glocker, Nikos Komodakis, Nikos Paragios.

- **Rapid Deformable Part Model Detection**: in [27] we introduce a Branch-and-Bound technique which efficiently finds the most promising configuration of a pictorial structure model given an image. The fastest previously known techniques are linear in the image size; our technique has a best-case complexity that is logarithmic in the image size. When evaluated on standard datasets (Pascal benchmark) our technique gives a 5- to 15-fold speedup. Moreover, when evaluated in the multi-object detection problem our technique’s complexity scales sublinearly also in the number of objects, resulting in 20- to 100-fold speedups when evaluated with 20 object categories.
• **Segmentation with Deformable Graph-based Priors:** in [22] we have introduced a novel formulation to address deformable segmentation using graph-based priors while being able to handle partial-correspondences. Segmentation was formulated as a matching task, where candidate correspondences were determined using boosting, and the assignment problem was solved using MAP inference constrained by a graph-based deformable prior. The notion of missing/erroneous correspondences was introduced in the process leading to state-of-the-art results once compared with prior art in the field. The same prior was used in the context of the segmentation of tagging MR heart images [37]. The main contribution of this paper was the exact estimation of the region-based probability likelihood within a pair-wise MRF through the use of Stokes theorem and integral images.

• **Deformable Model-based 3D reconstruction:** in [23] we introduce a model-based optimization approach to the 3D reconstruction of Femur images using a small set of low-dose X-Ray images. We use a parametric deformable model of the Femur surface and fit it to the acquired data by optimizing its parameters. We incorporate in our optimization criterion multiple aspects of the problem, namely the 3D surface-to-2D plane projection, region-based statistics, and edge-based terms. Our evaluation includes both in vitro and in vivo experiments, where our method is shown to yield promising results, while alleviating the need for time demanding, manual annotations.

• **Pose-invariant Higher Order Graph-based Priors:** in [36] we have introduced a novel method for 3D model inference from 2D images in the absence of camera pose parameters. The method exploits higher (fourth) order priors, which alleviate the need of the estimation of the camera parameters. Furthermore, the proposed formulation couples 3D model inference with 2D correspondences and results on a single shot solution for both problems in the absence of knowledge of the observer internal and external parameters.

### 6.3. Fusion/Registration

**Participants:** Stavros Alchatzidis, Nicolas Honnorat, Fabrice Michel, Aristeidis Sotiras, Chaohui Wang, Alex Bronstein, Michael Bronstein, Christos Davatzikos, Ben Glocker, Nikos Komodakis, Yangming Ou, Dimitris Samaras, Regis Vaillant, Yun Zeng, Nikos Paragios.

• **Intrinsic Dense 3D Surface Matching:** in [38] a probabilistic tracking framework for registering two 3D shape that relies on accurate correspondences between all points across the two frames was proposed. The definition of the matching cost is done using the “uniformization” theory that is combined with regularization terms that enforce spatial and temporal motion consistencies, into a maximum a posteriori (MAP) problem which we approximate using a Markov Random Field (MRF).

• **Optimal Linear Registration:** in [26] we proposed a novel formulation to address linear registration of volumetric images (translation, rotation and scale) that guarantees the optimality of the obtained solution. This was achieved through the approximation of the volumetric data using a sparse representation and the expression of the registration criterion in the form of a difference of convex functions. Cutting plain algorithms in the high-dimensional space were used to provide the optimal solution of the registration problem.

• **Quasi-real Time Registration:** in [21] we proposed a novel message-passing based optimization method to for pair-wise Markov Random Fields models and their applications in medical imaging and computer vision. Such a method was integrated to the deformable registration paradigm introduced in [12]. Such an optimization framework was combined with efficient use of modern architectures (Graphics Processing Units) leading to a speed up of at least one order of magnitude with respect to [12] making quasi real-time deformable registration feasible.

• **Metric Learning:** in [31] we extend prior work on similarity sensitive hashing to address multimodal 3D registration. The method consists of combining invariant to translation/rotation/scale features defined at the Gabor space with a machine learning/boosting method that aims to projection corresponding visual patterns to binary vectors with minimal Hamming distance while maximizing the distance between no corresponding samples.
• **Symmetric Deformable Fusion**: in [9] a novel graph-based formulation combining image and geometric terms was proposed for deformable registration. The method aimed at constraining iconic registration using a set of landmark correspondences that are sparse, do not inherit redundancy and are symmetric. The central idea was to simultaneously deform the target and the source image using two symmetric flows such that the similarity criterion is reaching its lowest potential. This was achieved through the use of composite symmetric deformation fields. This formulation was expressed as a graph-based optimization problem leading to promising experimental results.

• **Deformable registration of gene expression data**: in [28] the combined iconic/geometric registration framework introduced in [9] was extended to deal with gene expression data. Similarity Sensitive Hashing was used to establish costs for landmark correspondences, and a graph-based formulations with unknowns the deformation vectors was adopted for the objective function. Such an idea was extended to deal with combined segmentation/registration approach through an atlas in [29] where subdivision surfaces were considered to represent the deformation grid.

6.4. **Physiological Modeling & Spatio-Temporal Analysis**

**Participants**: Nicolas Honnorat, Sarah Parisot, Stephane Chemouny, Hugues Dufaut, Regis Vaillant, Nikos Paragios.

• **Low Gliomas Brain Map**: in [33] we introduce a graph-based modeling approach towards spatial position interpretation of low gliomas brain tumors. This was achieved through unsupervised clustering from exemplars, where spatial and geometric proximity of tumors were used to determine the strength connectivity of a graph. Towards automatic estimation of the lowest rank graph that is able to express the observed variation of tumors, an LP problem was solved that determines automatically the number of clusters and their centers while associating the training exemplars with them. Promising results that are well aligned with observations from neuro-sciences demonstrate the potentials of the proposed formulation.

• **Coupled Iconic/Geometric Spatio-temporal Segmentation**: in [25] we have introduced a combined elongated structures segmentation/tracking approach that was based on a two-layer graphical model. The image layer was exploiting the visual space and was seeking to minimize a data-driven cost while the geometric layers was seeking to establish temporal correspondences of the deforming structure. These two layers were coupled through a common set of variables acting on the deformation of the control points representing the elongated structure. Guide-wire segmentation [24] and tracking in low signal-to-noise ratio interventional images demonstrated the extreme potentials of our approach.
5. New Results

5.1. New results: geometric control

A first set of new results concerns sub-Riemannian geometry.

- In [3] we continued the study of almost-Riemannian structures, which are rank-varying sub-Riemannian structures locally generated by a number of vector fields equal to the dimension of the ambient manifold. In particular, two-dimensional almost-Riemannian structures are generalized Riemannian structures on surfaces for which local orthonormal frames are Lie bracket generating pair of vector fields that can become collinear. We considered the Carnot–Carathéodory distance canonically associated with an almost-Riemannian structure and studied the problem of Lipschitz equivalence between two such distances on a given compact oriented surface. We analyzed the generic case, allowing in particular for the presence of tangency points, i.e., points where two generators of the distribution and their Lie bracket are linearly dependent. The main result of the paper provides a characterization of the Lipschitz equivalence class of an almost-Riemannian distance in terms of a labeled graph associated with it.

- In [1] we studied nilpotent 2-step, corank 2 sub-Riemannian metrics. Such metrics naturally appear as nilpotent approximations of general sub-Riemannian ones. We exhibited optimal syntheses for these problems. It turns out that in general the cut time is not equal to the first conjugate time but has a simple explicit expression. As a byproduct of this study we proved some smoothness properties of the spherical Hausdorff measure in the case of a generic 6-dimensional, 2-step corank 2 sub-Riemannian metric.

- In [12] we started from the remark that in Carnot–Carathéodory spaces the class of 1-rectifiable sets does not contain smooth non-horizontal curves. We were looking for a new definition of rectifiable sets including non-horizontal curves. We introduced, for any metric space, a new class of curves, called continuously metric differentiable of degree \( k \), which are Hölder but not Lipschitz continuous when \( k > 1 \). Replacing Lipschitz curves by this kind of curves we defined \((H^k, 1)\)-rectifiable sets and showed a density result generalizing the corresponding one in Euclidean geometry. This theorem has been obtained as a consequence of computations of Hausdorff measures along curves, for which we gave an integral formula. In particular, we showed that both spherical and standard Hausdorff measures along curves coincide with a class of dimensioned lengths and are related with an interpolation complexity, for which estimates have already been obtained in Carnot–Carathéodory spaces.

A class of problems for which tracking and motion planning is crucial, is given by the control of unmanned aerial vehicles (UAV). In order to develop improved planning tasks that take into account payload requirements, optimal costs and obstacles avoidance (or no flight zones), it is important to develop reliable and flexible simulators. One such simulator for a UAV ground control station is proposed in [9]. The research focuses on the connection between the UAV trajectories and its sensors. Our proposal includes a module-based description of the architecture of the simulator and is based on a nonlinear model of a fixed wing aircraft.

5.2. New results: quantum control

New results have been obtained for the control of the bilinear Schrödinger equation, with two different approaches.
• In [2] we proved an approximate controllability result by finite-dimensional methods, considering the Galerkin approximations. The approach improves the technique that we developed in [40]. The result requires less restrictive non-resonance hypotheses on the spectrum of the uncontrolled Schrödinger operator than those already known. The control operator is not required to be bounded and we are able to extend the controllability result to the density matrices. The proof is based on fine controllability properties of the finite-dimensional Galerkin approximations and allows to get estimates for the $L^1$ norm of the control. The general controllability result is applied to the problem of controlling the rotation of a bipolar rigid molecule confined on a plane by means of two orthogonal external fields.

• In [4] we presented a constructive method to control the bilinear Schrödinger equation via two controls. The method is based on adiabatic theory and works if the spectrum of the Hamiltonian admits conical eigenvalue intersections. We provided sharp estimates of the relation between the error and the controllability time. We also showed that for a Hamiltonian of the kind $-\Delta + V_0(x) + u_1V_1(x) + u_2V_2(x)$ on a domain of $\mathbb{R}^n$ the eigenvalue intersections are conical generically with respect to $V_0, V_1, V_2$.

5.3. New results: neurophysiology

We gave new contributions to the developing theory of human locomotion modeled through optimal control problems. In this paradigm, the trajectories are assumed to be solutions of an optimal control problem whose cost has to be determined.

• The purpose of [6] has been to analyze the class of optimal control problems defined in this way. We proved strong convergence of their solutions, on the one hand for perturbations of the initial and final points (stability), and on the other hand for perturbations of the cost (robustness).

• In [5] we discussed the modeling of both the dynamical system and the cost to be minimized, and we analyzed the corresponding optimal synthesis. The main results describe the asymptotic behavior of the optimal trajectories as the target point goes to infinity.

In [10] we studied the model of geometry of vision due to Petitot, Citti and Sarti [81]. One of the main features of this model is that the primary visual cortex V1 lifts an image from $\mathbb{R}^2$ to the bundle of directions of the plane. Neurons are grouped into orientation columns, each of them corresponding to a point of this bundle. In this model a corrupted image is reconstructed by minimizing the energy necessary for the activation of the orientation columns corresponding to regions in which the image is corrupted. The minimization process intrinsically defines an hypoelliptic heat equation on the bundle of directions of the plane. In the original model, directions are considered both with and without orientation giving rise respectively to a problem on the group of rototranslations of the plane $SE(2)$ or on the projective tangent bundle of the plane. We provided a mathematical proof of several important facts for this model. We first proved that the model is mathematically consistent only if directions are considered without orientation. We then proved that the convolution of a $L^2(\mathbb{R}^2, \mathbb{R})$ function (e.g. an image) with a 2D Gaussian is generically a Morse function. This fact is important since the lift of Morse functions to the projective tangent bundle of the plane is defined on a smooth manifold. We then provided the explicit expression of the hypoelliptic heat kernel on the projective tangent bundle of the plane in terms of Mathieu functions. Finally, we presented the main ideas of an algorithm which allows to perform image reconstruction on real non-academic images. The algorithm is massively parallelizable and needs no information on where the image is corrupted.

5.4. New results: switched systems

New results on switched systems have been obtained in three directions:

• Discrete-time systems. In [14] we dealt with the stability properties of linear discrete-time switched systems with polytopic sets of modes. The most classical and viable way of studying the uniform asymptotic stability of such a system is to check for the existence of a quadratic Lyapunov function. It
is known from the literature that letting the Lyapunov function depend on the time-varying switching parameter improves the chance that a quadratic Lyapunov function exists. The contribution of [14] is twofold. We first proved that under a non-degeneracy assumption the dependence on the switching function can be actually assumed to be linear with no prejudice on the effectiveness of the method. Moreover, we showed that no gain is obtained even if we allow the Lyapunov function to depend on the time. Second, we introduced the notion of eventual accessible sets and we showed that, in the degenerate case, it leads to a relaxation of the LMI conditions to check stability of switched linear systems. As a consequence, equivalence between different notions of quadratic stability can still be established under an additional assumption but, in general, allowing the Lyapunov function to depend on time leads to less conservative LMI conditions, as we explicitly showed through an example. We also discussed the case where the variation of the switching parameter is bounded by a prescribed constant between two subsequent times.

- **Continuous-time systems subject to persistent-excitation.** In [11] we studied linear control systems for which the controlled part can be switched off by a signal subject to a persistent excitation condition. We were interested in the stabilization problem of this system by a linear state feedback and we positively answered a question asked in [41], proving the following: Assume that the class of persistently exciting signals is restricted to those which are $M$-Lipschitzian, where $M > 0$ is a positive constant. Then, given any $C > 0$, there exists a linear state feedback depending on the class of signals under consideration (but not an individual signal) so that the rate of exponential decay of the time-varying system associated with any signal is greater than $C$.

- **Infinite-dimensional continuous-time systems.** In [13] we partially extended the analysis of finite-dimensional systems subject to persistently exciting signals to the case of systems driven by PDEs. More precisely, we studied the asymptotic stability of a dissipative evolution in a Hilbert space subject to intermittent damping. We observed that, even if the intermittence satisfies a persistent excitation condition, if the Hilbert space is infinite-dimensional then the system needs not being asymptotically stable (not even in the weak sense). Exponential stability is recovered under a generalized observability inequality, allowing for time-domains that are not intervals. Weak asymptotic stability is obtained under a similarly generalized unique continuation principle. Strong asymptotic stability is proved for intermittences that do not necessarily satisfy some persistent excitation condition, evaluating their total contribution to the decay of the trajectories of the damped system. Our results are discussed using the example of the wave equation and the linear Schrödinger equation.
6. New Results

6.1. Mesh Generation and Geometry Processing

6.1.1. Isotropic 2D Quadrangle Meshing with Size and Orientation Control

Participants: Pierre Alliez, Bertrand Pellenard.

In collaboration with Jean-Marie Morvan from University of Lyon.

We propose an approach for automatically generating isotropic 2D quadrangle meshes from arbitrary domains with a fine control over sizing and orientation of the elements. At the heart of our algorithm is an optimization procedure that, from a coarse initial tiling of the 2D domain, enforces each of the desirable mesh quality criteria (size, shape, orientation, degree, regularity) one at a time, in an order designed not to undo previous enhancements. Our experiments demonstrate how well our resulting quadrangle meshes conform to a wide range of input sizing and orientation fields. [31].

Figure 1. The algorithm takes as input a 2D domain, a sizing field and a cross field (not shown). It then operates on a triangle background mesh: The initialization clusters background mesh triangles so that the tiling roughly meets the size and shape criteria; A relaxation then improves the tiling for shape and orientation while preserving size; A conforming relaxation improves the degree of the tiles and the regularity of the tiling; A series of local parameterizations further improves the degrees and regularity; Barycentric subdivision generates a pure quadrangle mesh; Smoothing finally improves the shape of the quadrangles. We depict the conformance both to the sizing and to the cross field.
6.1.2. An Optimal Transport Approach to Robust Reconstruction and Simplification of 2D Shapes

Participants: Pierre Alliez, David Cohen-Steiner.

In collaboration with Fernando de Goes and Mathieu Desbrun from Caltech.

We propose a robust 2D shape reconstruction and simplification algorithm which takes as input a defect-laden point set with noise and outliers. We introduce an optimal-transport driven approach where the input point set, considered as a sum of Dirac measures, is approximated by a simplicial complex considered as a sum of uniform measures on 0- and 1-simplices. A fine-to-coarse scheme is devised to construct the resulting simplicial complex through greedy decimation of a Delaunay triangulation of the input point set. Our method performs well on a variety of examples ranging from line drawings to grayscale images, with or without noise, features, and boundaries. [25].

![Figure 2. Robustness to noise and outliers.](image)

Figure 2. Robustness to noise and outliers. The input shape (3K points) has sharp corners subtending small angles as well as boundaries. Our reconstruction is perfect for a noise-free input (left); as noise is added (middle, 2% and 2.5% of bounding box), the output degrades gracefully, still capturing most of the sharp angles; even after adding 4K or 4.5K outliers and 2% of noise (right), the reconstruction remains of quality, although artifacts start appearing in this regime.

6.1.3. Anisotropic Delaunay Mesh Generation

Participants: Jean-Daniel Boissonnat, Mariette Yvinec.

In collaboration with Camille Wormser from Google.

Anisotropic meshes are triangulations of a given domain in the plane or in higher dimensions, with elements elongated along prescribed directions. Anisotropic triangulations are known to be well suited for interpolation of functions or solving PDEs. Assuming that the anisotropic shape requirements for mesh elements are given through a metric field varying over the domain, we propose a new approach to anisotropic mesh generation, relying on the notion of anisotropic Delaunay meshes. An anisotropic Delaunay mesh is defined as a mesh in which the star of each vertex \( v \) consists of simplices that are Delaunay for the metric associated to vertex \( v \). This definition works in any dimension and allows to define a simple refinement algorithm. The algorithm takes as input a domain and a metric field and provides, after completion, an anisotropic mesh whose elements are shaped according to the metric field. [46]

6.1.4. Triangulating Smooth Submanifolds with Light Scaffolding

Participants: Jean-Daniel Boissonnat, Arijit Ghosh.
We propose an algorithm to sample and mesh a $k$-submanifold $M$ of positive reach embedded in $\mathbb{R}^d$. The algorithm first constructs a crude sample of $M$. It then refines the sample according to a prescribed parameter $\epsilon$, and builds a mesh that approximates $M$. Differently from most algorithms that have been developed for meshing surfaces of $\mathbb{R}^3$, the refinement phase does not rely on a subdivision of $\mathbb{R}^d$ (such as a grid or a triangulation of the sample points) since the size of such scaffoldings depends exponentially on the ambient dimension $d$. Instead, we only compute local stars consisting of $k$-dimensional simplices around each sample point. By refining the sample, we can ensure that all stars become coherent leading to a $k$-dimensional triangulated manifold $\tilde{M}$. The algorithm uses only simple numerical operations. We show that the size of the sample is $O(\epsilon^{-k})$ and that $\tilde{M}$ is a good triangulation of $M$. More specifically, we show that $M$ and $\tilde{M}$ are isotopic, that their Hausdorff distance is $O(\epsilon^2)$ and that the maximum angle between their tangent bundles is $O(\epsilon)$. The asymptotic complexity of the algorithm is $T(\epsilon) = O(\epsilon^{-k^2-k})$ (for fixed $M$, $d$ and $k$).

6.2. Topological and Geometric Inference

6.2.1. Metric graph reconstruction from noisy data

Participants: Frédéric Chazal, Marc Glisse.

In collaboration with Mridul Aanjaneya, Daniel Chen, Leonidas J. Guibas and Dmitriy Morozov.

Many real-world data sets can be viewed as noisy samples of special types of metric spaces called metric graphs. Building on the notions of correspondence and Gromov-Hausdorff distance in metric geometry, we describe a model for such data sets as an approximation of an underlying metric graph. We present a novel algorithm that takes as an input such a data set, and outputs the underlying metric graph with guarantees. We also implement the algorithm, and evaluate its performance on a variety of real world data sets.

6.2.2. Persistence-Based Clustering in Riemannian Manifolds

Participants: Frédéric Chazal, Steve Oudot.

In collaboration with Leonidas J. Guibas and Primoz Skraba.
We introduce a clustering scheme that combines a mode-seeking phase with a cluster merging phase in the corresponding density map. While mode detection is done by a standard graph-based hill-climbing scheme, the novelty of our approach resides in its use of topological persistence to guide the merging of clusters. Our algorithm provides additional feedback in the form of a set of points in the plane, called a persistence diagram (PD), which provably reflects the prominences of the modes of the density. In practice, this feedback enables the user to choose relevant parameter values, so that under mild sampling conditions the algorithm will output the correct number of clusters, a notion that can be made formally sound within persistence theory.

The algorithm only requires rough estimates of the density at the data points, and knowledge of (approximate) pairwise distances between them. It is therefore applicable in any metric space. Meanwhile, its complexity remains practical: although the size of the input distance matrix may be up to quadratic in the number of data points, a careful implementation only uses a linear amount of memory and takes barely more time to run than to read through the input. [29].

6.2.3. Data-driven trajectory smoothing

**Participant:** Frédéric Chazal.

*In collaboration with Daniel Chen, Leonidas J. Guibas, Xiaoye Jiang and Christian Sommer*

Motivated by the increasing availability of large collections of noisy GPS traces, we present a new data-driven framework for smoothing trajectory data. The framework, which can be viewed of as a generalization of the classical moving average technique, naturally leads to efficient algorithms for various smoothing objectives. We analyze an algorithm based on this framework and provide connections to previous smoothing techniques. We implement a variation of the algorithm to smooth an entire collection of trajectories and show that it performs well on both synthetic data and massive collections of GPS traces. [28].

6.2.4. A Weighted \( k \)-Nearest Neighbor Density Estimate for Geometric Inference

**Participants:** Frédéric Chazal, David Cohen-Steiner.
Motivated by a broad range of potential applications in topological and geometric inference, we introduce a weighted version of the $k$-nearest neighbor density estimate. Various pointwise consistency results of this estimate are established. We present a general central limit theorem under the lightest possible conditions. In addition, a strong approximation result is obtained and the choice of the optimal set of weights is discussed.

In particular, the classical $k$-nearest neighbor estimate is not optimal in a sense described in the manuscript. The proposed method has been implemented to recover level sets in both simulated and real-life data. [12].

6.2.5. Deconvolution for the Wasserstein metric and geometric inference

**Participants:** Frédéric Chazal, Claire Caillerie.

Recently, [17], [13] have defined a distance function to measures to answer geometric inference problems in a probabilistic setting. According to their result, the topological properties of a shape can be recovered by using the distance to a known measure $\nu$, if $\nu$ is close enough to a measure $\mu$ concentrated on this shape. Here, close enough means that the Wasserstein distance $W_2$ between $\mu$ and $\nu$ is sufficiently small. Given a point cloud, a natural candidate for $\nu$ is the empirical measure $\mu_n$. Nevertheless, in many situations the data points are not located on the geometric shape but in the neighborhood of it, and $\mu_n$ can be too far from $\mu$. In a deconvolution framework, we consider a slight modification of the classical kernel deconvolution estimator, and we give a consistency result and rates of convergence for this estimator. Some simulated experiments illustrate the deconvolution method and its application to geometric inference on various shapes and with various noise distributions. [14].

6.2.6. Manifold Reconstruction Using Tangential Delaunay Complexes

**Participants:** Jean-Daniel Boissonnat, Arijit Ghosh.

We give a new provably correct algorithm to reconstruct a $k$-dimensional manifold embedded in $d$-dimensional Euclidean space [44]. The input to our algorithm is a point sample coming from an unknown manifold. Our approach is based on two main ideas: the notion of tangential Delaunay complex and the technique of sliver removal by weighting the sample points. Differently from previous methods, we do not construct any subdivision of the $d$-dimensional ambient space. As a result, the running time of our algorithm depends only linearly on the extrinsic dimension $d$ while it depends quadratically on the size of the input sample, and exponentially on the intrinsic dimension $k$. This is the first certified algorithm for manifold reconstruction whose complexity depends linearly on the ambient dimension. We also prove that for a dense enough sample the output of our algorithm is ambient isotopic to the manifold and a close geometric approximation of the manifold.

6.2.7. Equating the witness and restricted Delaunay complexes

**Participants:** Jean-Daniel Boissonnat, Ramsay Dyer, Arijit Ghosh, Steve Oudot.

It is a well-known fact that the restricted Delaunay and witness complexes may differ when the landmark and witness sets are located on submanifolds of Rd of dimension 3 or more. Currently, the only known way of overcoming this issue consists of building some crude superset of the witness complex, and applying a greedy sliver exudation technique on this superset. Unfortunately, the construction time of the superset depends exponentially on the ambient dimension, which makes the witness complex based approach to manifold reconstruction impractical. This work [43] provides an analysis of the reasons why the restricted Delaunay and witness complexes fail to include each other. From this a new set of conditions naturally arises under which the two complexes are equal.

6.2.8. Reconstructing 3D compact sets

**Participant:** David Cohen-Steiner.

In collaboration with Frédéric Cazals.
Reconstructing a 3D shape from sample points is a central problem faced in medical applications, reverse engineering, natural sciences, cultural heritage projects, etc. While these applications motivated intense research on 3D surface reconstruction, the problem of reconstructing more general shapes hardly received any attention. This paper develops a reconstruction algorithm changing the 3D reconstruction paradigm as follows.

First, the algorithm handles general shapes i.e. compact sets as opposed to surfaces. Under mild assumptions on the sampling of the compact set, the reconstruction is proved to be correct in terms of homotopy type. Second, the algorithm does not output a single reconstruction but a nested sequence of plausible reconstructions. Third, the algorithm accommodates topological persistence so as to select the most stable features only. Finally, in case of reconstruction failure, it allows the identification of under-sampled areas, so as to possibly fix the sampling.

These key features are illustrated by experimental results on challenging datasets (see Figure 5), and should prove instrumental in enhancing the processing of such datasets in the aforementioned applications. [16].

6.3. Data Structures and Robust Geometric Computation

6.3.1. Explicit array-based compact data structures for triangulations

Participant: Olivier Devillers.

In collaboration with Luca Castelli Aleardi (LIX, Palaiseau).

We consider the problem of designing space efficient solutions for representing triangle meshes. Our main result is a new explicit data structure for compactly representing planar triangulations: if one is allowed to permute input vertices, then a triangulation with \( n \) vertices requires at most \( 4n \) references (\( 5n \) references if vertex permutations are not allowed). Our solution combines existing techniques from mesh encoding with a novel use of minimal Schnyder woods. Our approach extends to higher genus triangulations and could be
applied to other families of meshes (such as quadrangular or polygonal meshes). As far as we know, our solution provides the most parsimonious data structures for triangulations, allowing constant time navigation in the worst case. Our data structures require linear construction time, and all space bounds hold in the worst case. We have implemented and tested our results, and experiments confirm the practical interest of compact data structures [47], [35].

6.3.2. Hyperbolic Delaunay triangulations and Voronoi diagrams made practical

Participants: Mikhail Bogdanov, Olivier Devillers, Monique Teillaud.

We show how to compute Delaunay triangulations and Voronoi diagrams of a set of points in hyperbolic space in a very simple way. The algorithm is implemented in an exact and efficient way [34] (see Figure 6).

6.4. Applications

6.4.1. Study of the cosmic web

Participant: Monique Teillaud.

In collaboration with many coauthors: members of the OrbiCG Associate Team (Section 8.3.1.3), Herbert Edelsbrunner (IST Austria, Duke University, and Geomagic Inc.), and others

We introduce a new descriptor of the weblike pattern in the distribution of galaxies and matter: the scale dependent Betti numbers which formalize the topological information content of the cosmic mass distribution (see Figure 7). While the Betti numbers do not fully quantify topology, they extend the information beyond conventional cosmological studies of topology in terms of genus and Euler characteristic used in earlier analyses of cosmological models. The richer information content of Betti numbers goes along with the availability of fast algorithms to compute them. When measured as a function of scale they provide a “Betti signature” for a point distribution that is a sensitive yet robust discriminator of structure. The signature is highly effective in revealing differences in structure arising in different cosmological models, and is exploited towards distinguishing between different dark energy models and may likewise be used to trace primordial non-Gaussianities. In this study we demonstrate the potential of Betti numbers by studying their behaviour in simulations of cosmologies differing in the nature of their dark energy [48], [41]. This work uses previous results obtained in GEOMETRICA [49], [50].
Figure 7. Four α-shapes of a Voronoi filament model realization. It concerns a sample of 200000 particles in a periodic box of 50 h⁻¹Mpc size with 8 Voronoi cells. From top left to bottom right:
\[ \alpha = 0.5 \times 10^4, 1.0 \times 10^4, 2 \times 10^4 \text{ and } 4 \times 10^4. \]
6.5. Software

6.5.1. CGAL

Two major new releases of CGAL, versions 3.8 and 3.9, have been made available in 2011. These releases contain the following new features, involving GEOMETRICA researchers:

- **Generator.** In release 3.8, the package Generator has been extended to provide various point set generators in dimensions higher than 3. It can generate random point sets in/on a sphere, in a cube, and points on a grid [40].

- **Spatial sorting.** Spatial sorting allows to order a set of points to improve the efficiency of incremental randomized algorithms. The spatial sorting package was existing in previous releases, and has been extended to dimensions higher than 3 in release 3.9 [39].

- **3D Mesh Generation.** The mesh generation package was introduced in CGAL 3.5. From release CGAL 3.6, the package offers, after Delaunay refinement, an optional optimization step to either improve the global mesh quality or get rid of slivers. Release CGAL 3.7 includes an interactive demo based on Qt and the code has been optimized for efficiency. Release 3.8 and further [38] offer the possibility to preserve sharp features such as creases and corners when provided in the description of the input domain.

The new release also contains new packages implemented by our CGAL partners and improvements to some existing packages: a detailed list can be found on the CGAL web site.

Two one-week CGAL developers meetings take place each year. The last one, organized in September at INRIA Sophia Antipolis by Mariette Yvinec, gathered 20 participants.
GRAND-LARGE Project-Team

6. New Results

6.1. Communication avoiding algorithms for linear algebra

Participants: Laura Grigori, Simplice Donfack, Amal Khabou, Mathias Jacquelin, Sophie Moufawad.

The focus of this research is on the design of efficient parallel algorithms for solving problems in numerical linear algebra, as solving very large sets of linear equations and large least squares problems, often with millions of rows and columns. These problems arise in many numerical simulations, and solving them is very time consuming.

This research focuses on developing new algorithms for linear algebra problems, that minimize the required communication, in terms of both latency and bandwidth. We have introduced in 2008 two communication avoiding algorithms for computing the LU and QR factorizations, that we refer to as CALU and CAQR (joint work with J. Demmel and M. Hoemmen from U.C. Berkeley, J. Langou from C.U. Denver, and H. Xiang then at INRIA) [6], [9]. Since then, we have also designed a communication avoiding algorithm for rank revealing QR. In addition, we have also extended theoretical lower bounds to sparse Cholesky factorization and identified algorithms that attain these bounds and so minimize communication. The communication avoiding algorithms are now studied by several other groups, including groups at INRIA, and they start being implemented and being available in public libraries as ScaLAPACK.

During 2011, our research has focused on a study of the stability of communication avoiding LU factorization and on its implementation on multicore machines. In [20] we focus on numerical properties of CALU. To decrease the communication required in the LU factorization, CALU uses a new pivoting strategy, referred to as tournament pivoting, that may lead to a different row permutation than the classic LU factorization with partial pivoting. We have further investigated the numerical stability of CALU. The reason to consider CALU is that it does an optimal amount of communication, and asymptotically less than Gaussian elimination with partial pivoting (GEPP), and so will be much faster on platforms where communication is expensive, as shown in previous work. We show that the Schur complement obtained after each step of performing CALU on a matrix A is the same as the Schur complement obtained after performing GEPP on a larger matrix whose entries are the same as the entries of A (sometimes slightly perturbed) and zeros. More generally, the entire CALU process is equivalent to GEPP on a large, but very sparse matrix, formed by entries of A and zeros. Hence we expect that CALU will behave as GEPP and it will be also very stable in practice. In addition, extensive experiments on random matrices and a set of special matrices show that CALU is stable in practice. The upper bound on the growth factor of CALU is worse than of GEPP. However, there are Wilkinson like-matrices for which GEPP has exponential growth factor, but not CALU, and vice-versa.

We present experimental results for random matrices and for a set of special matrices, including sparse matrices, for binary tree based and flat-tree-based CALU. We discuss both the stability of the LU factorization and of the linear solver, in terms of pivot growth and backward errors. The results show that in practice CALU is stable. We present the backward errors measured three ways: by \( \| PA - LU \| / \| A \| \), by the normwise backward error \( \| Ax - b \| / (\| A \| \| x \| + \| b \| ) \), and by the componentwise backward error (after iterative refinement in working precision). For random matrices, all CALU’s backward errors were at most 1.9x larger than GEPP’s backward errors. We also tested "special" matrices, including known difficult examples: (1) The ratios of \( \| PA - LU \| / \| A \| \) were at most 1 in over 69% of cases (i.e. CALU was at least as stable as GEPP), and always 1.5 or smaller, except for one ratio of 4.3, in which case both backward errors were much smaller than \( 2^{-53} \) = machine epsilon. (2) The ratios of normwise backward errors were at most 1 in over 53% of cases, and always 1.5 or smaller, except for 5 ratios ranging up to 26, in which cases all backward errors were much smaller than \( 4 \times \) machine epsilon. (3) The ratios of componentwise backward errors were at most 1 in over 52% of cases, and always 3.2 or smaller, except for one ratio of 8.3.
In [30] we design a scheduling algorithm for efficiently executing CALU on multicore architectures. We focus on a tunable scheduling strategy that maintains load balance across cores while also maintaining data locality and low dequeue overhead. To achieve this, we use a strategy that combines static and dynamic scheduling. This approach was shown to be successful on regular mesh computations by V. Kale and B. Gropp. This tunable scheduling strategy allows us to flexibly control the percentage of tasks that can be scheduled dynamically; this gives a knob to control load balancing so that it occurs only at the point in computation when the benefits it provides outweighs the costs it induces. On NUMA machines where remote memory access is costly, the percentage of work scheduled dynamically should be small enough to avoid excessive cache misses, but large enough to keep the cores busy during idle times in the static part.

In this work, we show the effectiveness of this method in the context of already highly-optimized dense matrix factorizations. Our prior work on multi-threaded CALU was based on dynamic scheduling. The algorithm performed well on tall and skinny matrices, but became less scalable on square matrices with increasing numbers of processors. We show that the usage of this scheduling in communication avoiding dense factorization leads to significant performance gains. On a 48 core AMD Opteron NUMA machine, our experiments show that we can achieve up to 64% improvement over a version of CALU that uses fully dynamic scheduling, and up to 30% improvement over the version of CALU that uses fully static scheduling. On a 16-core Intel Xeon machine, our hybrid static/dynamic scheduling approach is up to 8% faster than the version of CALU that uses a fully static scheduling or fully dynamic scheduling. Our algorithm leads to speedups over the corresponding routines for computing LU factorization in well known libraries. On the 48 core AMD NUMA machine, our best implementation is up to 110% faster than MKL, while on the 16 core Intel Xeon machine, it is up to 82% faster than MKL. Our approach also shows significant speedups compared with PLASMA on both of these systems.

6.2. Preconditioning techniques for solving large systems of equations

Participants: Laura Grigori, Riadh Fezzanni, Sophie Moufawad.

A different direction of research is related to preconditioning large sparse linear systems of equations. This research is performed in the context of ANR PETALh project (2011-2012), which follows the ANR PETAL project (2008-2009). It is conducted in collaboration with Frederic Nataf from University Paris 6.

Several highly used preconditioners are for example the incomplete LU factorizations and Schwarz based approaches as used in domain decomposition. Most of these preconditioners are known to have scalability problems. The number of iterations can increase significantly when the size of the problem increases or when the number of independent domains is increased. This is often due to the presence of several low frequency modes that hinder the convergence of the iterative method. To address this problem, we study a different class of preconditioners, called direction preserving or filtering preconditioners. These preconditioners have the property of being identical to the input matrix on a given filtering vector. A judicious choice of the vector allows to alleviate the effect of low frequency modes on the convergence.

We consider in particular two classes of preconditioners. The first preconditioner is an incomplete decomposition that satisfies the filtering property [11]. The nested preconditioner has the same property for a specific vector of all ones. However the construction is different and takes advantage of a nested structure of the input matrix. The previous research on these methods considered only matrices arising from the discretization of PDEs on structured grids, where the matrix has a block tridiagonal structure. This structure imposes a sequential computation of the preconditioner and it is not suitable for the more general case of unstructured grids. Hence, while very efficient, the usage of these preconditioners was very limited. At the beginning of this research we have obtained several theoretical results for these methods that demonstrate their numerical behavior and convergence properties for cases arising from the discretization of PDEs on structured grids [11]. But the main result is the development of a generalized method [48], [46] that has two important properties: it allows the filtering property to be satisfied for any input matrix; the matrix can be reordered such that its computation is highly parallel. Experimental results show that the method is very efficient for certain classes of matrices, and shows good scalability results in terms of both problem size and number of processors.
6.3. Microwave Data Analysis for petaScale computers

Participants: Laura Grigori, Mikolaj Szydlarski, Meisam Sharify.

In [47] we describe an scalable algorithm for computing an inverse spherical harmonic transform suitable for cluster of multiple CPU-GPUs. We base our implementation on hybrid programming combining MPI and CUDA. We focus our attention on the two major sequential steps involved in the transforms computation, retaining the efficient parallel framework of the original code. We detail optimization techniques used to enhance the performance of the OpenMP/CUDA-based code and compare them with those implemented in the public domain parallel package, S2HAT.

We also present performance comparisons of the multi GPU version and a hybrid, MPI/OpenMP version of the same transform. We find that one NVIDIA Tesla S1070 can accelerate overall execution time of the SHT by as much as 3 times with respect to the MPI/OpenMP version executed on one quad-core processor (Intel Nehalem 2.93 GHz) and, owing to very good scalability of both versions, 128 Tesla cards perform as good as 256 twelve-core processor (AMD Opteron 2.1 GHz).

The work presented here has been performed in the context of the Cosmic Microwave Background simulations and analysis. However, we expect that the developed software will be of more general interest and applicability.

6.4. Innovative linear system solvers for hybrid multicore/GPU architectures

Participant: Marc Baboulin.

The advent of new processor architectures (e.g. multicore, GPUs) requires the rethinking of most of the scientific applications and innovative methods must be proposed in order to take full advantage of current supercomputers [12].

To accelerate linear algebra solvers on current parallel machines, we introduced in public domain libraries a class of solvers based on statistical techniques. A first application concerns the solution of a square linear systems $Ax = b$. We study a random transformation of $A$ that enables us to avoid pivoting and then to reduce the amount of communication [54]. Numerical experiments show that this randomization can be performed at a very affordable computational price while providing us with a satisfying accuracy when compared to partial pivoting. This random transformation called Partial Random Butterfly Transformation (PRBT) is optimized in terms of data storage and flops count. In the solver that we developed, PRBT combined with LU factorization with no pivoting take advantage of the latest generation of hybrid multicore/GPU machines and outperform existing factorization routines from current parallel library MAGMA.

A second application is related to solving symmetric indefinite systems via $LDL^T$ factorization for which there was no existing parallel implementation in the dense library ScaLAPACK. We developed an efficient and innovative parallel tiled algorithm for solving symmetric indefinite systems on multicore architectures [59] & [1]. This solver avoids pivoting by using a multiplicative preconditioning based on symmetric randomization. This randomization prevents the communication overhead due to pivoting, is computationally inexpensive and requires very little storage. Following randomization, a tiled LDLT factorization is used that reduces synchronization by using static or dynamic scheduling. We compare Gflop/s performance of our solver with other types of factorizations on a current multicore machine and we provide tests on accuracy using LAPACK test cases.

6.5. MILEPOST GCC: machine learning enabled self-tuning compiler

Participant: Grigori Fursin [correspondent].

Tuning compiler optimizations for rapidly evolving hardware makes porting and extending an optimizing compiler for each new platform extremely challenging. Iterative optimization is a popular approach to adapting programs to a new architecture automatically using feedback-directed compilation. However, the large number of evaluations required for each program has prevented iterative compilation from widespread take-up in production compilers. Machine learning has been proposed to tune optimizations across programs systematically but is currently limited to a few transformations, long training phases and critically lacks publicly released, stable tools.
Our approach is to develop a modular, extensible, self-tuning optimization infrastructure to automatically learn the best optimizations across multiple programs and architectures based on the correlation between program features, run-time behavior and optimizations. In this paper we describe MILEPOST GCC, the first publicly-available open-source machine learning-based compiler. It consists of an Interactive Compilation Interface (ICI) and plugins to extract program features and exchange optimization data with the cTuning.org open public repository. It automatically adapts the internal optimization heuristic at function-level granularity to improve execution time, code size and compilation time of a new program on a given architecture. Part of the MILEPOST technology together with low-level ICI-inspired plugin framework is now included in the mainline GCC.

We developed machine learning plugins based on probabilistic and transductive approaches to predict good combinations of optimizations. Our preliminary experimental results show that it is possible to automatically reduce the execution time of individual MiBench programs on various machines from GRID5000, some by more than a factor of 2, while also improving compilation time and code size. We also present a realistic multi-objective optimization scenario for Berkeley DB library using MILEPOST GCC and improve execution time by approximately 17%, while reducing compilation time and code size by 12% and 7% respectively on Intel Xeon processor.

6.6. Loop Transformations: Convexity, Pruning and Optimization

Participant: Cédric Bastoul.

High-level loop transformations are a key instrument in mapping computational kernels to effectively exploit resources in modern processor architectures. However, determining appropriate compositions of loop transformations to achieve this remains a significantly challenging task; current compilers may achieve significantly lower performance than hand-optimized programs. To address this fundamental challenge, we first present a convex characterization of all distinct, semantics-preserving, multidimensional affine transformations. We then bring together algebraic, algorithmic, and performance analysis results to design a tractable optimization algorithm over this highly expressive space. The framework has been implemented and validated experimentally on a representative set of benchmarks run on state-of-the-art multi-core platforms.

6.7. Exact algorithm for the l1-compressive sensing problem using a modified Dantzig-Wolfe method

Participants: Alexandre Borghi, Jerome Darbon, Sylvain Peyronnet.

In this work, we consider the l1-Compressive Sensing problem and presents an efficient algorithm that computes an exact solution. The idea consists in reformulating the problem such that it yields a modified Dantzig-Wolfe decomposition that allows to efficiently apply all standard simplex pivoting rules. Experimental results show the superiority of our approach compared to standard linear programming methods.

6.8. Supple: a flexible probabilistic data dissemination protocol for wireless sensor networks

Participants: Aline Carneiro Viana, Thomas Hérault, Thomas LArgillier, Sylvain Peyronnet, Fatiha Zaidi.

We propose a flexible proactive data dissemination approach for data gathering in self-organized Wireless Sensor Networks (WSN). Our protocol Supple, effectively distributes and stores monitored data in WSNs such that it can be later sent to or retrieved by a sink. Supple empowers sensors with the ability to make on the fly forwarding and data storing decisions and relies on flexible and self-organizing selection criteria, which can follow any predefined distribution law. Using formal analysis and simulation, we show that Supple is effective in selecting storing nodes that respect the predefined distribution criterion with low overhead and limited network knowledge.
6.9. Non-self-stabilizing and self-stabilizing gathering in networks of mobile agents—the notion of speed

Participants: Joffroy Beauquier, Janna Burman, Julien Clément, Shay Kutten.

In the population protocol model, each agent is represented by a finite state machine. Agents are anonymous and supposed to move in an asynchronous way. When two agents come into range of each other (“meet”), they can exchange information. One of the vast variety of motivating examples to the population protocols model is ZebraNet. ZebraNet is a habitat monitoring application where sensors are attached to zebras and collect biometric data (e.g. heart rate, body temperature) and information about their behavior and migration patterns (via GPS). The population protocol model is, in some sense, related to cloud computing and to networks characterized by asynchrony, large scale, the possibility of failures, in the agents as well as in the communications, with the constraint that each agent is resource limited.

In order to extend the computation power and efficiency of the population protocol model, various extensions were suggested. Our contribution is an extension of the population protocol model that introduces the notion of “speed”, in order to capture the fact that the mobile agents move at different speeds and/or have different communication ranges and/or move according to different patterns and/or visit different places with different frequencies. Intuitively, fast agents which carry sensors with big communication ranges communicate with other agents more frequently than other agents do. This notion is formalized by allocating a cover time, cv, to each mobile agent v. cv is the minimum number of events in the whole system that occur before agent v meets every other agent at least once. As a fundamental example, we have considered the basic problem of gathering information that is distributed among anonymous mobile agents and where the number of agents is unknown. Each mobile agent owns a sensed input value and the goal is to communicate the values (as a multi-set, one value per mobile agent) to a fixed non-mobile base station (BS), with no duplicates or losses.

Gathering is a building block for many monitoring applications in networks of mobile agents. For example, a solution to this problem can solve a transaction commit/abort task in MANETs, if the input values of agents are votes (and the number of agents is known to BS). Moreover, the gathering problem can be viewed as a formulation of the routing problem in Disruption Tolerant Networks.

We gave different solutions to the gathering in the model of mobile agents with speed and we proved that one of them is optimal.

6.10. Making Population Protocols Self-stabilizing

Participants: Joffroy Beauquier, Janna Burman, Shay Kutten, Brigitte Rozoy.

As stated in the previous paragraph, the application domains of the population protocol model are asynchronous large scale networks, in which failures are possible and must be taken into account. This work concerns failures and namely the technique of self-stabilization for tolerating them.

Developing self-stabilizing solutions (and proving them) is considered to be more challenging and complicated than developing classical solutions, where a proper initialization of the variables can be assumed. This remark holds for a large variety of models and hence, to ease the task of the developers, some automatic techniques have been proposed to transform programs into self-stabilizing ones.

We have proposed such a transformer for algorithms in the population protocol model introduced for dealing with resource-limited mobile agents. The model we consider is a variation of the original one in that there is a non mobile agent, the base station, and that the communication characteristics (e.g. moving speed, communication radius) of the agents are considered through the notion of cover time.

The automatic transformer takes as an input an algorithm solving a static problem and outputs a self-stabilizing solution for the same problem. To the best of our knowledge, it is the first time that such a transformer for self-stabilization is presented in the framework of population protocols. We prove that the transformer we propose is correct and we make the complexity analysis of the stabilization time.
6.11. Self-stabilizing synchronization in population protocols with cover times
Participants: Joffroy Beauquier, Janna Burman, Shay Kutten, Brigitte Rozoy.

Synchronization is widely considered as an important service in distributed systems which may simplify protocol design. Phase clock is a general synchronization tool that provides a form of a logical time. We have developed a self-stabilizing phase clock algorithm suited to the model of population protocols with cover time. We have shown that a phase clock is impossible in the model with only constant-state agents. Hence, we assumed an existence of resource unlimited agent - the base station. The clock size and duration of each phase of the proposed phase clock tool are adjustable by the user. We provided application examples of this tool and demonstrate how it can simplify the design of protocols. In particular, it yields a solution to Group Mutual Exclusion problem.

6.12. Impossibility of consensus for population protocol with cover times
Participants: Joffroy Beauquier, Janna Burman.

We have extended the impossibility result for asynchronous consensus of Fischer, Lynch and Paterson (FLP) to the asynchronous model of population protocols with cover times. We noted that the proof of FLP does not apply. Indeed, the key lemma stating that two successive factors in an execution, involving disjoint subsets of agents, commute, is no longer true, because of the cover time property. Then we developed a completely different approach and we proved that there is no general solution to consensus for population protocols with cover times, even if there is a single possible crash. We noted that this impossibility result also applies to randomized asynchronous consensus, contrary to what happens in the classical message-passing or shared memory communication models, in which the problem is solvable inside some bounds on the number of faulty processes. Then, for circumventing these impossibility results, we introduced the phase clock oracle and the S oracle, and we shown how they allow to design solutions.

6.13. Routing and synchronization in large scale networks of very cheap mobile sensors
Participants: Joffroy Beauquier, Brigitte Rozoy.

In a next future, large networks of very cheap mobile sensors will be deployed for various applications, going from wild life preserving or environmental monitoring up to medical or industrial system control. Each sensor will cost only a few euros, allowing a large scale deployment. They will have only a few bit of memory, no identifier, weak capacities of computation and communication, no real time clock and will be prone to failures. Moreover such networks will be fundamentally dynamic. The goal of this subject is to develop the basic protocols and algorithms for rudimentary distributed systems for such networks. The studied problems are basic ones, like data collection, synchronization (phase clock, mutual exclusion, group mutual exclusion), fault tolerance (consensus), automatic transformers, always in a context of possible failures. A well known model has already been proposed for such networks, the population protocol model. In this model, each sensor is represented by a finite state machine. Sensors are anonymous and move in an asynchronous way. When two sensors come into range of each other (“meet”), they can exchange information. One of the vast variety of motivating examples for this model is ZebraNet. ZebraNet is a habitat monitoring application in which sensors are attached to zebras in order to collect biometric data (e.g., heart rate, body temperature) and information about their behavior and migration patterns. Each pair of zebras meets from time to time. During such meetings (events), ZebraNet’s agents (zebras’ attached sensors) exchange data. Each agent stores its own sensor data as well as data of other sensors that were in range in the past. They upload data to a base station whenever it is nearby. It was shown that the set of applications that can be solved in the original model of population protocols is rather limited. Other models (such as some models of Delay/Disruption-Tolerant Networks - DTNs), where each node maintains links and connections even to nodes it may interact with only intermittently, do not seem to suit networks with small memory agents and a very large (and unknown) set of anonymous agents. That is why we enhance the model of population protocols by introducing a notion...
of "speed". We try to capture the fact that the mobile agents move at different speeds and/or have different communication ranges and/or move according to different patterns and/or visit different places with different frequencies. Intuitively, fast agents which carry sensors with large communication ranges communicate with other agents more frequently than other agents do. This notion is formalized by the notion of cover time for each agent. The cover time of an agent is the unknown number of events (pairwise meetings) in the whole system that occur (during any execution interval) before agent v meets every other agent at least once. The model we propose is somehow validated by some recent statistical results, obtained from empirical data sets regarding human or animal mobility. An important consequence of our approach is that the analytic complexity of the protocols designed in this model is possible, independently of any simulation or experimentation. For instance, we consider the fundamental problem of gathering different pieces of information, each sensed by a different anonymous mobile agent, and where the number of agents is unknown. The goal is to communicate the sensed values (as a multi-set, one value per mobile agent) to a base station, with no duplicates or losses. Gathering is a building block for many monitoring applications in networks of mobile agents. Moreover, the gathering problem can be viewed as a special case of the routing problem in DTNs, in which there is only one destination, the base station. Then we are able to compute the complexity of solutions we propose, as well as those of solutions used in experimental projects (like ZebraNet), and to compare them. The algorithms we present are self-stabilizing. Such algorithms have the important property of operating correctly regardless of their initial state (except for some bounded period). In practice, self-stabilizing algorithms adjust themselves automatically to any changes or corruptions of the network components (excluding the algorithm’s code). These changes are assumed to cease for some sufficiently long period. Self-stabilization is considered for two reasons. First, mobile agents are generally fragile, subject to failures and hard to initialize. Second, systems of mobile agents are by essence dynamic, some agents leave the system while new ones are introduced. Self-stabilization is a well adapted framework for dealing with such situations.

6.14. Self-Stabilizing Control Infrastructure for HPC

Participants: Thomas Hérault, Camille Coti.

High performance computing platforms are becoming larger, leading to scalability and fault-tolerance issues for both applications and runtime environments (RTE) dedicated to run on such machines. After being deployed, usually following a spanning tree, a RTE needs to build its own communication infrastructure to manage and monitor the tasks of parallel applications. Previous works have demonstrated that the Binomial Graph topology (BMG) is a good candidate as a communication infrastructure for supporting scalable and fault-tolerant RTE.

In this work, we presented and analyzed a self-stabilizing algorithm to transform the underlying communication infrastructure provided by the launching service (usually a tree, due to its scalability during launch time) into a BMG, and maintain it in spite of failures. We demonstrated that this algorithm is scalable, tolerates transient failures, and adapts itself to topology changes.

The algorithms are scalable, in the sense that all process memory, number of established communication links, and size of messages are logarithmic with the number of elements in the system. The number of synchronous rounds to build the system is also logarithmic, and the number of asynchronous rounds in the worst case is square logarithmic with the number of elements in the system. Moreover, the self-stabilizing property of the algorithms presented induce fault-tolerance and self-adaptivity. Performance evaluation based on simulations predicts a fast convergence time (1/33s for 64K nodes), exhibiting the promising properties of such self-stabilizing approach.

We pursue this work by implementing and evaluating the algorithms in the STCI runtime environment to validate the theoretical results.

6.15. Large Scale Peer to Peer Performance Evaluations

Participant: Serge Petiton.
6.15.1. Large Scale Grid Computing

Recent progress has made possible to construct high performance distributed computing environments, such as computational grids and cluster of clusters, which provide access to large scale heterogeneous computational resources. Exploration of novel algorithms and evaluation of performance is a strategic research for the future of computational grid scientific computing for many important applications [88]. We adapted [68] an explicit restarted Lanczos algorithm on a world-wide heterogeneous grid platform. This method computes one or few eigenpairs of a large sparse real symmetric matrix. We take the specificities of computational resources into account and deal with communications over the Internet by means of techniques such as out-of-core and data persistence. We also show that a restarted algorithm and the combination of several paradigms of parallelism are interesting in this context. We perform many experimentations using several parameters related to the Lanczos method and the configuration of the platform. Depending on the number of computed Ritz eigenpairs, the results underline how critical the choice of the dimension of the working subspace is. Moreover, the size of platform has to be scaled to the order of the eigenproblem because of communications over the Internet.

6.15.2. High Performance Cluster Computing

Grid computing focuses on making use of a very large amount of resources from a large-scale computing environment. It intends to deliver high-performance computing over distributed platforms for computation and data-intensive applications. We propose [99] an effective parallel hybrid asynchronous method to solve large sparse linear systems by the use of a Grid Computing platform Grid5000. This hybrid method combines a parallel GMRES(m) (Generalized Minimum RESidual) algorithm with the Least Square method that needs some eigenvalues obtained from a parallel Arnoldi algorithm. All of these algorithms run on the different processors of the platform Grid5000. Grid5000, a 5000 CPUs nation-wide infrastructure for research in Grid computing, is designed to provide a scientific tool for computing. We discuss the performances of this hybrid method deployed on Grid5000, and compare these performances with those on the IBM SP series supercomputers.

6.15.3. Large Scale Power aware Computing

Energy conservation is a dynamic topic of research in High Performance Computing and Cluster Computing. Power-aware computing for heterogeneous world-wide Grid is a new track of research. We have studied and evaluated the impact of the heterogeneity of the computing nodes of a Grid platform on the energy consumption. We propose to take advantage of the slack-time caused by the heterogeneity in order to save energy with no significant loss of performance by using Dynamic Voltage Scaling (DVS) in a distributed eigensolver [69]. We show that using DVS only during the slack-time does not penalize the performances but it does not provide significant energy savings. If DVS is applied to all the execution, we get important global and local energy savings (respectively up to 9% and 20%) without a significant rise of the wall-clock times.

6.16. High Performance Linear Algebra on the Grid

Participants: Thomas Hérault, Camille Coti.

Previous studies have reported that common dense linear algebra operations do not achieve speed up by using multiple geographical sites of a computational grid. Because such operations are the building blocks of most scientific applications, conventional supercomputers are still strongly predominant in high-performance computing and the use of grids for speeding up large-scale scientific problems is limited to applications exhibiting parallelism at a higher level.

In this work, we have identified two performance bottlenecks in the distributed memory algorithms implemented in ScALAPACK, a state-of-the-art dense linear algebra library. First, because ScALAPACK assumes a homogeneous communication network, the implementations of ScALAPACK algorithms lack locality in their communication pattern. Second, the number of messages sent in the ScALAPACK algorithms is significantly greater than other algorithms that trade flops for communication.
This year, we presented a new approach for computing a QR factorization one of the main dense linear algebra kernels of tall and skinny matrices in a grid computing environment that overcomes these two bottlenecks. Our contribution is to articulate a recently proposed algorithm (Communication Avoiding QR) with a topology-aware middleware (QCG-OMPI) in order to confine intensive communications (ScalAPACK calls) within the different geographical sites.

An experimental study conducted on the Grid5000 platform shows that the resulting performance increases linearly with the number of geographical sites on large-scale problems (and is in particular consistently higher than ScalAPACKs).

6.17. Emulation of Volatile Systems

Participants: Thomas Largillier, Benjamin Quetier, Sylvain Peyronnet, Thomas Hérault, Franck Cappello.

In the process of developing grid applications, people need to often evaluate the robustness of their work. Two common approaches are simulation, where one can evaluate his software and predict behaviors under conditions usually unachievable in a laboratory experiment, and experimentation, where the actual application is launched on an actual grid. However, simulation could ignore unpredictable behaviors due to the abstraction done and experimentation does not guarantee a controlled and reproducible environment, and simulation often introduces a high level of abstraction that makes the discovery and study of unexpected, but real, behaviors a rare event.

In this work, we proposed an emulation platform for parallel and distributed systems including grids where both the machines and the network are virtualized at a low level. The use of virtual machines allows us to test highly accurate failure injection since we can destroy virtual machines, and network virtualization provides low-level network emulation. Failure accuracy is a criteria that evaluates how realistic a fault is. The accuracy of our framework has been evaluated through a set of micro benchmarks and a very stable P2P system called Pastry.

We are in the process of developing a fault injection tool to work with the platform. It will be an extension of the work started in the tool Fail. The interest of this work is that using Xen virtual machines will allow to model strong adversaries since it is possible to have virtual machines with shared memory. These adversaries will be stronger since they will be able to use global fault injection strategies.

6.18. Exascale Systems

Participant: Franck Cappello.

Over the last 20 years, the open-source community has provided more and more software on which the world’s high-performance computing systems depend for performance and productivity. The community has invested millions of dollars and years of effort to build key components. Although the investments in these separate software elements have been tremendously valuable, a great deal of productivity has also been lost because of the lack of planning, coordination, and key integration of technologies necessary to make them work together smoothly and efficiently, both within individual petascale systems and between different systems. A repository gatekeeper and an email discussion list can coordinate open-source development within a single project, but there is no global mechanism working across the community to identify critical holes in the overall software environment, spot opportunities for beneficial integration, or specify requirements for more careful coordination. It seems clear that this completely uncoordinated development model will not provide the software needed to support the unprecedented parallelism required for peta/exascale computation on millions of cores, or the flexibility required to exploit new hardware models and features, such as transactional memory, speculative execution, and GPUs. We presented a rational promoting that the community must work together to prepare for the challenges of exascale computing, ultimately combining their efforts in a coordinated International Exascale Software Project.
Over the past few years resilience has became a major issue for high-performance computing (HPC) systems, in particular in the perspective of large petascale systems and future exascale systems. These systems will typically gather from half a million to several millions of central processing unit (CPU) cores running up to a billion threads. From the current knowledge and observations of existing large systems, it is anticipated that exascale systems will experience various kind of faults many times per day. It is also anticipated that the current approach for resilience, which relies on automatic or application level checkpoint/restart, will not work because the time for checkpointing and restarting will exceed the mean time to failure of a full system. This set of projections leaves the community of fault tolerance for HPC systems with a difficult challenge: finding new approaches, which are possibly radically disruptive, to run applications until their normal termination, despite the essentially unstable nature of exascale systems. Yet, the community has only five to six years to solve the problem. In order to start addressing this challenge, we synthesized the motivations, observations and research issues considered as determinant of several complimentary experts of HPC in applications, programming models, distributed systems and system management.

As a first step to address the resilience challenge, we conducted a comprehensive study of the state of the art. The emergence of petascale systems and the promise of future exascale systems have reinvigorated the community interest in how to manage failures in such systems and ensure that large applications, lasting several hours or tens of hours, are completed successfully. Most of the existing results for several key mechanisms associated with fault tolerance in high-performance computing (HPC) platforms follow the rollback-recovery approach. Over the last decade, these mechanisms have received a lot of attention from the community with different levels of success. Unfortunately, despite their high degree of optimization, existing approaches do not fit well with the challenging evolutions of large-scale systems. There is room and even a need for new approaches. Opportunities may come from different origins: diskless checkpointing, algorithmic-based fault tolerance, proactive operation, speculative execution, software transactional memory, forward recovery, etc.

We provided the following contributions: (1) we summarize and analyze the existing results concerning the failures in large-scale computers and point out the urgent need for drastic improvements or disruptive approaches for fault tolerance in these systems; (2) we sketch most of the known opportunities and analyze their associated limitations; (3) we extract and express the challenges that the HPC community will have to face for addressing the stringent issue of failures in HPC systems.
6. New Results

6.1. Massive mobile dense wireless networks

Participants: Cédric Adjih, Aline Carneiro Viana, Emmanuel Baccelli, Philippe Jacquet, Pascale Minet, Paul Mühlethaler, Yasser Toor.

6.1.1. Executive summary

Scaling properties of mobile ad hoc networks lead to an increase of global capacity when the network density increases or when the packets can be stored for a while in mobile nodes instead of being immediately retransmitted.

Gupta and Kumar have shown in 2000 that the transport capacity per node in a multihop ad hoc network decreases in \( 1 / \sqrt{N \log N} \), \( N \) being the number of nodes in the network. Therefore the global capacity of the network increases in \( \sqrt{N} / \sqrt{\log N} \). This is a surprising result since in wired networks a collection of nodes connected to a single communication resource has a transport capacity that just remains constant (i.e., the average per node capacity decreases in \( 1/N \)).

Therefore adding space to a multihop wireless network increases the capacity: this is the space capacity paradox.

When nodes randomly move, it turns to be more advantageous to store packets for a while on mobile routers instead of forwarding them immediately like hot potatoes. When the mobile router moves closer to the destination, then it can deliver packets on a much smaller number of hops. Of course the delivery delay is much longer, but the network capacity also increases by slowing non-urgent packets. This is the time capacity paradox: by slowing packets, nodes mobility increases network capacity. This was hinted the first time by Grossglauser and Tse in 2002.

The great challenge is to find the good protocol and tunings that allow to adjust the delivery delay from zero to infinity in order to get a continuous increase in capacity. The challenge is two-sided: one has to keep the delivery delay between reasonable bounds and one has to consider realistic mobility models.

Existing protocols for Mobile Ad Hoc Networks (MANET) are highly efficient in routing data between mobile nodes that belong to the same connected component (cf. the protocols which have received the RFC status by the MANET group of IETF). What about a disconnected network where source and destination may be located in two different connected components? In this case usual routing protocols drop packet due to host unreachable as no end-to-end route exists at that time.

A simple idea is to allow the router that has no available route to the destination to keep the packet in buffer until the conditions become more appropriate for forwarding. The forwarding conditions will change because of mobility: the router can move closer to the destination so that they belong to the same connected component and the packet can be delivered.

Indeed, the network may be continuously partitioned due to high mobility, and the traditional approach to allow a mobile node to wait for the network to be fully connected (i.e., form a unique component) or to wait to be in range of the destination may lead to unacceptable delays. Furthermore, concrete applications, such as Defence and Disaster-Relief, cannot always rely on such assumptions.

Nevertheless, even if the communicating nodes may never be within the same connected component, it is important to observe that a “communication path” may be available through time using intermediate nodes that are temporarily within reach of each other while moving, hence making such networks viable for critical applications. Depending on the nature of the environment, these networks are now commonly referred as Intermittently Connected MANET and Delay Tolerant Networks.
In between stands the problem of the fully connected network that forms a single connected component, but for which maintaining full knowledge of the topology would simply make the network collapse under its huge control traffic. In fact this is the main problem that wireless network engineering has to face, in most experiments the generation of control traffic is the main source of disruption.

6.1.2. Scientific achievements

6.1.2.1. Scaling and spatial capacity in non uniform wireless networks

We found a more precise instance of Gupta-Kumar result by using a simple but realistic network model based on slotted ALOHA with Poisson traffic. It turns out that when the traffic density increases then the average node neighborhood area shrinks so that the average encircled traffic load remains constant with an analytical expression.

In their original model Gupta and Kumar assume that the traffic density is constant, which is far from realistic. However we have derived similar generalized results when the traffic density is not uniform. In this case, the heavier is the local traffic, the smaller are the local neighborhood and the larger is the number of hops needed to cross the congested region. Therefore the shortest paths (in hop number as computed by OLSR) will have a natural tendency to avoid congested areas. The path tend to follow trajectory that have analogy in non linear optic with variable indices.

6.1.2.2. Time capacity and node mobility

We have defined a protocol that takes advantage of node mobility in a general way. In short the packet stay with its host router as long as the latter does not evade too fast from its next hop (computed via a shortest path protocol such as OLSR). In the way we understand “too fast” stands the tuning parameters we discussed above. There is no need to have node geographical location and to physically measure motion vector, since everything can be done via the analysis of the dynamic of neighborhood intersections. We analytically derived performance evaluation under random walk mobility models. We plan to simulate the protocol in a real mobility scenario. This algorithm has application in Intelligent Transport System.

6.1.2.3. Overhead reduction in large networks

The first limitation of multihop wireless network is the size of the overhead per node that increases linearly with the size of the network. This is a huge improvement compared to classic internet protocols which have quadratic overhead increases. Nevertheless this still limit the network size to some thousands. We have analyzed the performance of OLSR with Fisheye feature that significantly reduce the overhead with respect to distance. In theory the overhead reduction allows to network size of several order of magnitude. Anyhow the tuning of the overhead attenuation with distance must be carefully done when the network is mobile, in order to avoid tracking failure. We showed that an overhead reduction within square root of the network size achieve this goal.

An alternative way to overhead reduction is ad hoc hierarchical routing and Distributed Hashing Table. Work has just begun in this area.

6.1.2.4. Coloring in wireless networks

Coloring is used in wireless networks to improve communication efficiency, mainly in terms of bandwidth, energy and possibly end-to-end delays. Nodes access the medium according to their color. It is the responsibility of the coloring algorithm to ensure that interfering nodes do not have the same color. First, we established complexity results about the h-hop coloring problem. Second, we focused on wireless sensor networks with grid topologies. We proposed the Vector-Based Coloring Method, denoted VCM, a new method that is able to provide an optimal periodic coloring for any radio transmission range and for any h-hop coloring, h>=1. Third, we also designed OSERENA "Optimized SchEduling RoutEr Node Activity", a distributed coloring algorithm optimized for dense wireless networks.

6.1.2.5. Complexity results about the h-hop coloring problem

In the paper we published at the WMNC 2011 conference, we define the h-hop node coloring problem, with h any positive integer, adapted to two types of applications in wireless networks. We specify both general mode for general applications and strategic mode for data gathering applications. We prove that the associated decision problem is NP-complete.
6.1.2.6. Grid coloring and the Vector-Based Coloring Method

In 2011, we also focused on wireless sensor networks with grid topologies. How does a coloring algorithm take advantage of the regularity of grid topology to provide an optimal periodic coloring, that is a coloring with the minimum number of colors? We propose the Vector-Based Coloring Method, denoted VCM, a new method that is able to provide an optimal periodic coloring for any radio transmission range and for any h-hop coloring, h\geq 1. This method consists in determining at which grid nodes a color can be reproduced without creating interferences between these nodes while minimizing the number of colors used. We compare the number of colors provided by VCM with the number of colors obtained by a distributed coloring algorithm with line and column priority assignments. We also provide bounds on the number of colors of optimal general colorings of the infinite grid, and show that periodic colorings (and thus VCM) are asymptotically optimal. Finally, we discuss the applicability of this method to a real wireless network.

6.1.2.7. Opportunistic routing

The model of wireless networks based on dynamic graph does not well assess the real processes in a wireless network. In particular the range of transmission can greatly vary between packets, the graph keeping only the average range. Opportunistic routing consists into taking advantage of temporary extension of the transmission range in order to gain several hops.

We have strong established theoretical performance limits in opportunistic routing. The limits are based on realistic interference scenarios in slotted Aloha. We have also investigated the impact of mobility on this theoretical limits.

We have designed an opportunistic routing protocol whose performance are within a small margin of the theoretical limits.

We have also conducted studies to support intelligent and adaptive forwarding, which allows a good trade-off between reliability and resource-efficiency. We then design a new protocol, called GrAnt, a new prediction-based forwarding protocol for complex and dynamic delay tolerant networks (DTNs). The proposed protocol uses the Ant Colony Optimization (ACO) metaheuristic with a greedy transition rule. This allows GrAnt to select the most promising forwarder nodes or allow for the exploitation of previously found good paths. The main motivation for using ACO is to take advantage of its population-based search and the rapid adaptation of its learning framework. Considering data from heuristic functions and pheromone concentration, the GrAnt protocol includes three modules: routing, scheduling, and buffer management.

6.1.2.8. Intermittent and delay tolerant networks

We consider the problem of routing in these networks, with the sole assumption that the speed of the node mobility is less than the speed of transmitting a packet to a neighbour. We compare this problem with sound propagation in liquid. We show that various pattern of mobility and network clustering can be described by a single parameter such as the information speed propagation.

We introduce new algorithms that route a packet toward a remote destination. The different algorithms vary depending on the buffering and the capacity capabilities of the network (i.e. if one or more copies of a packet can be sent and/or be kept). All algorithms are based on link aging rumors across connected components. The packet bounces from connected components to connected components, thanks to node mobility. We establish several analytical properties using an analogy with the sound propagation in liquid where molecules creates temporary connected components where sounds travel very fast.

Previous models assumed that the propagation of information path evolves like in a dynamic Erdos-Renyi graph leading to an epidemic flooding in O(log n) or O(1), n being the number of nodes in the network. We disprove the Erdos-Renyi model by showing via space-time considerations that the set of information path from a source to a destination is in fact much smaller than the path set in the Erdos-Renyi model. This lead to a much larger minimal delay in square root of n instead of log n. This correspond to a bounded maximal information propagation speed, whose estimate depends on the mobility model and the node density, and is root of multivariate explicit Bessel formulas.
Additionally, we have also considered the problem of data collection in global sensing and intermittently connected systems while avoiding the use of costly infrastructures (e.g., 3G). Motivated by the observation that node encounters are sufficient to build a connected relationship graph, we propose to take advantage of such inherent interactions to transform some mobile devices into delegates. We use then opportunistic delegation as a data traffic offload solution by investigating two main questions: (i) How to gain insights into social mobile networking scenarios?, (ii) How to utilize such insights to design solutions to alleviate overloaded 3G networks?. Our solution leverages usage of mobile applications requiring large data transfers by channeling the traffic to a few, socially important users in the network called VIP delegates. Mobile collectors need then only to meet delegates that, in turn, are responsible for gathering data from a subset of standard producers. We first investigate several delegation strategies based on the relative importance of nodes in their social interactions. Second, by considering a prediction strategy that estimates the likelihood of two nodes meeting each other, we investigate how the delegation strategies perform on predicted traces.

6.1.2.9. Network Coding

We study network coding for multi-hop wireless networks. We focus on the case of broadcasting where one source transmits information to all nodes in the network. Our goal is energy-efficient broadcast, that is, minimizing the total number of transmissions for broadcasting to the entire network. Note that this is a different problem for the classical problem of capacity maximization ; and assuming we are far from the network capacity limit, hence in fact, we could assume interference-free transmissions.

Our previous results, they had shown that network coding (and a simple coding strategy) was able to reach optimality for asymptotically large and dense networks, with asymptotically 100 % of the received transmissions being useful (innovative). We extended the results with the combined use of connected dominating sets and network coding: we were able to quantify (and bound) the benefits of network coding in networks where the area of the network stays fixed, and only the density increases.

We have proved that the performance of wireless random network coding are optimal in the following network model: the Erdos-Renyi random graph model and the unit disk random graph model. In particular we show in the Erdos-Renyi the network coding capacity rate outperform any Connected Dominating Set strategy by a factor of order log n. In the unit disk model we gain is larger than 60%. The result is based on the analysis of the connectivity stretch ratio of the random graphs. The connectivity stretch ratio is the ratio of the smallest degree over the connectivity number, and the connectivity stretch ratio tends to one in the two graph models.

6.1.3. Collaborations

- Professor Bernard Mans, Macquarrie University, Sydney, Australia,
- TREC INRIA team,
- Professors Anelise Munaretto and Myriam Regattieri Delgado from Federal Technological University of Parana (UTFPR), Brazil,
- CNRS researcher Marcelo Dias de Amorim, LIP6/UPMC, France,
- Mathias Boc, CEA LIST, France,
- Computer Science Department, Sapienza University of Rome, Italy,
- University of St. Andrews, UK.
- Professor Leila Saidane, ENSI, Tunisia.

6.2. New generation of OLSR, new services and protocols

Participants: Cédric Adjih, Aline Carneiro Viana, Emmanuel Baccelli, Thomas Clausen, Philippe Jacquet, Pascale Minet, Ichrak Amdouni, Ridha Soua, Erwan Livolant, Paul Mühlethaler, Yasser Toor.

6.2.1. Executive summary

The user of a mobile network very quickly experience problems with quality of service: links fade, connectivity disrupts, delays accumulate.
In a wireless network, the set of neighbors which with one node can communicate depends on transmission range, and numerous factors, and in addition the transmission range is often lower than the interference range (the range within which a node prevents correct transmissions of other nodes). Thus bandwidth reservation, a crucial step of quality of service, is an important and difficult problem.

The services and protocols that need careful adaptation are

- Connectivity continuity
- Bandwidth reservation
- Delay routing
- Connectivity control
- Autoconfiguration
- Security
- Energy efficiency
- Localization

The connectivity continuity is the most important problem. Trivial in the wired world where a link failure is a rare event, it becomes problematic in the mobile world where link failure caused by mobility are frequent and normal. The first experiments of mobile ad hoc networks with regular internet protocols miserably failed simply because either the protocol was to slow to recover link failure, or when tuned appropriately was generating such a huge overhead that the network collapsed under its own weight. A new generation of routing protocols has arised that allow a suitable control of connectivity in mobile networks. Among them the Optimized Link State Routing combines the optimization of overhead for mobile networks and the full internet legacy. It naturally provides path redundancy which accelerate link failure recovery.

The most important lesson that must be retained is that most of these optimization become NP complete, which is a significant complication compared to their counterpart in the classical wired world. The reason for the NP-completeness is two-sided: on one side the co-interferences make impossible an optimization link by link, on the other side, the large dispersion of performance measurement makes simple heuristic ineffective. As an example, routing with respect to shortest delay average does not guarantee smallest probability of high delay.

Since the bandwidth is scarce, any multimedia application such as video streaming is resource demanding. For example a TV broadcast that uses a mesh network will rapidly exhaust the bandwidth if all connections are point to point. In this case multicast protocols that allows to gather all these point to point connections in a single flow is a need.

There are two classes of multicast protocols: the tree based protocols and the network coding protocols. In the first class the protocols take advantage of the relatively small size of the recipient node set. One can show equivalent results of Gupta and Kumar scaling properties but in the multicast plan when the ratio of recipient versus network size is a fundamental parameter. When this ratio tends to one the performance naturally worsen.

When the recipient set is the whole network, one can apply the network coding scheme with random packet combination. In network coding the packets are no longer isolated: relay nodes makes linear combination of packets and transmitted mixed packets. In theory the performance of network coding is better than isolated packet multicast. In practice network coding is simpler to operate does not need topology management such as spanning trees or Connected Dominating Set. The reason for this is highly non intuitive, as if packet superposition was acting like state superposition in quantum mechanic, leading to non expected results.

Quality of service has become the central requirement that users expect from a network. High throughput, service continuity are critical issue for multimedia application over the wireless internet where the bandwidth is more scarce than in the wired world. A significant issue in the ad-hoc domain is that of the integrity of the network itself. Routing protocols allow, according to their specifications, any node to participate in the network - the assumption being that all nodes are behaving well and welcome. If that assumption fails - then the network may be subject to malicious nodes, and the integrity of the network fails. An important security service over mobile networks is to ensure that the integrity of the network is preserved even when attacks are launched against the integrity of the network.
6.2.2. Scientific achievements

6.2.2.1. Optimized Link State Routing (OLSR)

The routing protocol OLSR is universally known in the mobile wireless community (more than 475,000 hits on Google). It has numerous implementations and is used in many wireless networks. It is a proactive protocol with full internet legacy which is based on partial topology information exchange, that non the less provide optimal path with additive metrics (such as BGP/OSPF). It is an experimental RFC within IETF and soon will become a full standard under the name OLSRv2.

6.2.2.2. OSPF extension for wireless mesh networking

Long a near-future myth, ad hoc networks are now becoming a reality as a variety of wireless mesh networks are being deployed. Wireless mesh networks are a specific kind of ad hoc network, where terminals are essentially fixed. Even in such cases, which somewhat resembles usual networks, specific routing protocols have nevertheless to be employed, to cope with the characteristics of wireless, multi-hop communications. Such characteristics include scarce bandwidth over inherently unreliable, versatile, semi-broadcast links, and absence of a central authority in general. One of the main difficulties in this context is to cope with contradictory requirements such as, on one hand, dealing with bandwidth scarcity, which typically requires decreasing control traffic, while on the other hand, dealing with unreliable, versatile links which typically requires increasing control traffic. The two prominent routing protocols that have been developed for ad hoc networks and studied over the past decade, are the IETF standards AODV and OLSR. AODV is based on a reactive scheme (i.e. on-demand flooding to discover a path to a new destination), while OLSR is based on a proactive scheme, which is essentially an optimization of link state routing (i.e. pre-provisioning of paths to all possible destinations). OLSR is to date the most deployed such protocol, as it powers numerous wireless mesh community networks that currently flourish in various cities throughout Europe and North America. Based on this experience, the integration of ad hoc networking in the "standard" networking body is going further in several directions. One direction is the IEEE 802.11s standardization effort, which uses AODV and OLSR-derived algorithms to provide wireless mesh routing capabilities below IP. Another direction, spearheaded by the IETF, is the extension of IP routing standards such as OSPF to support ad hoc routing: in this realm we recently spun RFC 5449, as well as a series of academic publications on the subject. The idea behind extending OSPF to support ad hoc networks comes from a simple observation: OSPF is algorithmically quite similar to OLSR, as both are based on a proactive, link state approach. As on the other hand OSPF is a well-understood, widely deployed, industry-standard protocol, employing it to integrate ad hoc networks with existing infrastructure is considered by users as an easy migration path.

6.2.2.3. Multi-metric routing

Quality of service involves finding routes between two nodes in the network that satisfies a number of constraints. These constraints could be the requested bandwidth, the maximum delay, the minimum loss probability, the reliability of links, etc. This problem is NP-Complete because it combines additive metrics in the optimization problem. Hipercom proposed heuristics for finding routes that respect up to four metrics when calculating routes between source and destination. Another QoS issue is the creation of models that estimate the actual value of a metric. For example, computing the available bandwidth or the transfer delay on a link, etc. is very complex in a non-deterministic medium access such as WiFi. To resolve this problem, we developed a model for estimating the available bandwidth in a wireless network. This model is based on considering interfering links in cliques, after which we provide the maximum capacity that could be deployed in a clique. We may still enhance the model by adding a scaling factor to the clique equations in order to become more accurate when compared to real measurements.

In particular we have investigated the metric based on packet delay distribution. Since propagation delays between routers are negligible, most delays occur in queueing and medium access control processing. Contrary to previous common belief there is no need of network synchronization. The objective is to proactively determine the delay in absence of packet data traffic. The estimate of delay distribution is done via analytical method. In order to keep control on quality of service flows we use source routing forwarding options.
6.2.3. Collaborations

- Many contractual collaborations:
  - MoD (QoS, security, interconnection between the OLSR and OSPF routing domains),
  - Hitachi (Vehicular applications, OLSRv2),
- Non contractual:
  - BAE (OLSRv2),
  - Deutsche Telekom Labs/TU-Berlin, Germany,

6.3. Wireless Sensor Networks

Participants: Cédric Adjih, Aline Carneiro Viana, Emmanuel Baccelli, Thomas Clausen, Philippe Jacquet, Pascale Minet, Ichrak Amdouni, Ridha Soua, Erwan Livolant, Paul Mühlthaler, Yasser Toor.

6.3.1. Executive summary

In wireless sensor networks, we focus more particularly on:

- Spatial reuse of the bandwidth,
- Routing according to a cross-layering approach,
- Security,
- Energy efficiency,
- IPv6 support.

6.3.2. Scientific achievements

6.3.2.1. Cryptographic Protocols to Fight Sinkhole Attacks on Tree-based Routing in Wireless Sensor Networks

Wireless Sensor Networks (WSN) are penetrating more and more in our daily life. As a consequence, security has become an important matter for these networks. We introduce two new cryptographic protocols of different complexity and strength in limiting network degradation caused by sinkhole attacks on tree-based routing topologies in Wireless Sensor Networks (WSNs). The main goal of both protocols is to provide continuous operation by improving resilience against, rather than detection of, these attacks. The main benefit of providing resilience is that it allows operating (or graceful degradation) in the presence of attacks. Furthermore, while resilience mechanisms do not dismiss detection mechanisms, detection mechanisms often introduce more complexity and so, more weaknesses to the system, which might not justify their benefits. More specifically our two RESl<em>l</em>ient and Simple Topology-based reconfiguration protocols are: RESIST-1 and RESIST-0. RESIST-1 prevents a malicious node from modifying its advertised distance to the sink by more than one hop, while RESIST-0 does not allow such lying at the cost of additional complexity.

6.3.2.2. IPv6 Protocol suite for Sensor Networks

Wireless sensor networking is a key element of the Internet of Things (IoT), a substantial part of the billions of smart objects that are soon to blend into the global IP network, from actuators to home appliances, from smart meters, to smart dust. Sensor nodes are devices used for distributed and automated monitoring of various parameters such as temperature, movement, noise or radioactivity levels etc. Sensors are scattered with minimum planning with respect to their precise physical position (including the central role of the sink, if any), and the set of peers with which a sensor can directly communicate through its wireless interface may change rapidly over time due to asynchronous sleep mode strategies, fluctuations in the radio environment, device failure or mobility. Through its wireless interface, a sensor thus connects to a communication link with undetermined connectivity properties. Sensor networks are a challenge to current IP standards, since on the one hand these protocols were designed to work on wired links and on the other hand these protocols were designed to work on machines that do not have drastic constraints in terms of CPU, power capacities, and memory, as sensor nodes do. In consequence, several key standard protocols (including TCP, UDP, DHCP,
NDP, SLAAC, and OSPF) do not function correctly in this environment. Nevertheless, IPv6-based sensor networking is a viable long term goal because it would enable generic, large scale, seamless integration of millions of sensing devices using heterogeneous radio technologies, at a low cost, and in a future-proof manner. The Internet Engineering Task Force (IETF) is currently engaged into multiple efforts addressing the limitations of existing standards concerning wireless sensor IP networking. Some of the standards under construction aim at fitting IP formats, especially IPv6 formats, to direct wireless communications using low power radio technologies such as IEEE 802.15.4, which require IP format compression. Other standards in development aim at providing multi-hop wireless sensor communication with IPv6, which requires specific routing protocols, efforts in which we actively participate, prompting numerous joint publications with both industrial and academic partners.

6.3.2.3. Coloring in wireless sensor networks
Graph coloring is used in wireless networks to optimize network resources: bandwidth and energy. We focus on grid topologies that constitute regular topologies for large or dense wireless networks. We consider various transmission ranges and identify a color pattern that can be reproduced to color the whole grid with the optimal number of colors. We obtain an optimal periodic coloring of the grid for the considered transmission range. We then evaluate the performance of a 3-hop distributed coloring algorithm, called SERENA. Through simulation results, we highlight the impact of node priority assignment on the number of colors obtained for any network and grids in particular. We then compare these optimal results on grids with those obtained by SERENA and identify directions to improve SERENA.

6.3.2.4. Coloring algorithm optimized for dense wireless networks
In 2011, we also designed OSERENA "Optimized SchEduling RoutEr Node Activity", a distributed coloring algorithm optimized for dense wireless networks. Network density has an extremely reduced impact on the size of the messages exchanged to color the network. Furthermore, the number of colors used to color the network is not impacted by this optimization. We describe the properties of the algorithm and prove its correctness and termination. Simulation results point out the considerable gains in bandwidth.

6.3.2.5. Multichannel access in wireless sensor networks
In 2011 we started a research activity on multichannel access in wireless sensor networks. A state of the art has been published at the IFIP Wireless Days 2011 Conference.

6.3.3. Collaborations

- Many contractual collaborations:
  - Hitachi (Vehicular applications, OLSRv2),
  - OCARI2 project (industrial wireless sensor network, QoS, cross layer, energy efficiency, routing, node activity scheduling),
  - SAHARA project (wireless sensor network embedded in aircrafts),
  - STIC INRIA-Tunisian Universities: the team of Prof. Leila Saidane at ENSI (Performance improvement in a wireless sensor network),

- Non contractual:
  - BAE (OLSRv2),
  - Freie Universitaet (sensor networks, DHT),
  - Deutsche Telekom Labs/TU-Berlin, Germany,
  - University of Athens, Greece.

6.4. Vehicular and mobile applications
Participants: Cédric Adjih, Emmanuel Baccelli, Thomas Clausen, Philippe Jacquet, Pascale Minet, Paul Mühlethaler, Yasser Toor.
6.4.1. Executive summary

We have the following vision: in the future mobile internet and static internet will have their core deeply intricated. This means that mobile ad hoc networks will be attached to the core network, form extension and even be part of it. For example in disaster area, a wireless network could replace the destroyed infrastructure and help to the emergency operations.

With this perspective items such as Autoconfiguration, Security are of crucial importance. However there is a potential conflict between a large population of fixed nodes based on ancient protocol and a smaller but more dynamic population based on new protocols. In the integration both population must cooperate in an hybrid protocol.

The difficulty is to build protocols that are as dynamic and efficient as MANET protocols but can support the legacy of the old and heavy internet protocols. The challenge is nevertheless achievable, because the dynamic part of the network needs less frequent updates from the fixed part of the network. Moreover the natural abundance of resource in the fixed part of the network allows it to support the more frequent updates from the mobile part.

OLSR has been found to be the natural best candidate for this challenge since it gathers dynamic and optimization with internet legacy.

6.4.2. Scientific achievements

6.4.2.1. Military tactical networks

During year 2011, we conducted several expertises about industrial propoals dealing with OLSR use in military tactical networks.

6.4.2.2. Protocols for vehicular networks

We have achieved numerous studies and design of protocols for vehicular networks and more spesifically for V2V (Vehicle-to-Vehicle) network.

First we have studied the channel occupancy induced by the OLSR proactive routing protocol used in a linear Vehicular Ad hoc Network (VANET). Unlike previous studies, which usually use simulations to evaluate the overhead, we have proposed a simple analytical model to carry out this evaluation. Moreover, we did not evaluate the total overhead induced by the routing protocol as is usually proposed, but, for a given node, the channel occupation induced by the routing protocol.

We have studied flooding techniques for safety applications in VANETs. The typical scenario is the diffusion of an alert message after a car crash in a platoon of vehicles. The packet is diffused with the pure flooding, the multipoint relay (MPR) diffusion of OLSR and a geographic aware protocol. For OLSR we have introduced a variant (Robust-MPR) to improve the reliability. Different realistic scenarios were considered and various parameters such as vehicle density, and background traffic load were scrutinized. We have shown that the Robust-MPR and the geographic aware protocol satisfy the requirements of the safety applications while using considerably less overhead than pure flooding.

We have shown that the geographic aware protocols can be improved for the diffusion of an alert message by using opportunistic routing. We have designed OB-VAN (Opportunistic Broadcast for VANets ) a new protocol that uses this idea. One of the novelty of this protocol is the use of an active signalling technique in the acknowledgement procedure to select the best relay taking advantage of the reception pattern of each message. We have studied OB-VAN in a linear VANET and have shown that it outperforms the flooding for the delay and the amount of overhead. However the delivery ratio of OB-VAN may be insufficient for safety applications. This remark has led to the design of R-OB-VAN which is a reliable variant of OB-VAN. With extensive simulations, we have shown that R-OB-VAN maintains a high delivery ratio even in the presence of packet loss due to shadowing.
We have studied the performance of the Aloha scheme in linear VANETs. This analysis assumes a SINR (Signal over Interference plus Noise Ratio) based model. In this model, we have derived the probability of a successful transmission between two vehicles at a distance of R meters. We have also computed the mean throughput according to Shannon’s law. In these two models, we have optimized the two quantities directly linked to the achievable network throughput i.e., the mean packet progress and the density of transport.

Finally, we have studied the utilization of opportunistic routing and shown that this technique is also beneficial for point to point traffic. It decreases the delay and increases the throughput compared with shortest path first routing. Moreover, we have also shown that opportunistic routing for point to point traffic eases considerably the optimization of the MAC scheme e.g. the transmission probability for Aloha and the carrier sense threshold for CSMA.

**6.4.3. Collaboration**

We received support from MoD for this activity.
6. New Results

6.1. Interaction Techniques

Participants: Caroline Appert, Michel Beaudouin-Lafon [correspondant], David Bonnet, Anastasia Bezerianos, Olivier Chapuis, Guillaume Faure, Emilien Ghomi, Stéphane Huot, Mathieu Nancel, Wendy Mackay, Cyprien Pindat, Emmanuel Pietriga, Theophanis Tsandilas, Julie Wagner.

Acquiring a target, such as pointing to an icon, a button or a landmark on a digital map, is the most common action in today’s graphical user interfaces. We have continued our work to better understand this seemingly simple action and make it faster and more reliable. This year we have conducted theoretical work on small target [12] and more practical work with TorusDesktop [4].

Targets of only a few pixels are notoriously difficult to acquire. Despite many attempts at facilitating pointing, the reasons for this difficulty are poorly understood. We confirm a strong departure from Fitts’ Law for small target acquisition using a mouse and investigate three potential sources of problems: motor accuracy, legibility, and quantization. We find that quantization is not a problem, but both motor and visual sizes are limiting factors. This suggests that small targets should be magnified in both motor and visual space to facilitate pointing. Since performance degrades exponentially as targets get very small, we further advocate the exploration of uniform, target-agnostic magnification strategies. We also confirm Welford’s 1969 proposal that motor inaccuracy can be modeled by subtracting a “tremor constant” from target size. We argue for the adoption of this model, rather than Fitts’ law, when reflecting on small target acquisition.

With TorusDesktop [4], we revisited a pointing technique that allows to wrap the mouse cursor around screen edges in conventional desktop environments. Allowing the cursor to jump from one edge of the screen to the opposite one (i.e., turning the desktop into a torus) was already explored, but never studied empirically nor designed for everyday desktop usage. We have introduced a dead zone and an off-screen cursor feedback that ease the use of this technique. We also conducted three controlled experiments to refine the design and evaluate its performance. Our results suggest clear benefits in several conditions, but also some potential limitations due to users’ over-estimation of cursor wrapping advantages. An implementation of TorusDesktop for the Mac OS X desktop can be downloaded for free at http://insitu.lri.fr/TorusDesktop.

We continued our work on wall-sized displays, focusing on the study of high-level tasks such as pan-zoom navigation (Figure 10), that have received little attention. Indeed, while pointing on this type of display has been studied extensively, it remains unclear which techniques are best suited to perform multiscale navigation in these environments. Building upon empirical data gathered from studies of pan-and-zoom on desktop computers and studies of remote pointing, we identified three key factors for the design of mid-air pan-and-zoom techniques: uni- vs. bimanual interaction, linear vs. circular movements, and level of guidance to accomplish the gestures in mid-air. After an extensive phase of iterative design and pilot testing, we ran a controlled experiment aimed at better understanding the influence of these factors on task performance. This work received a best paper award at CHI 2011 [6].

On the opposite side, we have studied small displays such as the ones smartphones are equipped with. One major challenge with this type of device is to make the user able to interact in parallel with both the device and other artefacts in his environment (e.g., giving a phone call while holding a paper document). The Swiss Army Menu (SAM) [27] is a radial menu that enables a very large number of functions accessible via small thumb movements. The design of SAM relies on four different kinds of items, support for navigating in hierarchies of items and a control that requires only the thumb of the hand that holds the device. SAM can offer a set of functions so large that it would typically have required a number of widgets that could not have been displayed in a single viewport at the same time.
The different interaction techniques presented above are aimed at improving the control within a given representation. As a complement, we have also worked on improving the user’s experience by providing him with two (or more) representations of the data he is interacting with and ways to transition between these representations.

Glimpse [17] is a quick preview technique that smoothly transitions between document markup code (HTML, wiki markup or LaTeX) and its visual rendering (see Figure 11). It allows users to regularly check the code they are editing in-place, without leaving the text editor. Glimpse can complement classical preview windows by offering rapid overviews of code-to-document mappings and leaving more screen real-estate. As the technique smoothly show the links between the code and the rendered result, it can also help to learn how complex markup code will result in the final document (e.g., HTML tables or LaTeX formulae).

In collaboration with University of Toronto and OCAD University, we designed a novel visualization technique called ChronoLenses [14], aimed at supporting users in time-series visual exploration tasks. ChronoLenses perform on-the-fly transformation of the data points in their focus area, tightly integrating visual analysis with user actions, and enabling the progressive construction of advanced visual analysis pipelines, supporting tasks that require visualizing derived values, identifying correlations, or discovering anomalies beyond obvious outliers.

We further explored user understanding of data presented in Dual-Scale data charts, charts that incorporate two different data resolutions into one chart in order to emphasize data in regions of interest (focus) or to enable the comparison of data from distant regions (context) [13]. In collaboration with researchers from inria AVIZ, we presented a unified description of different Dual-Scale data charts, and we compared them in terms of user understanding using elementary graphical perception tasks, such as comparing lengths and distances. Our study suggests that cut-out charts which include colocated full context and focus are the best alternative, and that superimposed charts in which focus and context overlap on top of each other should be avoided.

6.2. Research Methods
InkSplore [19]. We conducted three studies with contemporary music composers at IRCAM. We found that even highly computer-literate composers use an iterative process that begins with expressing musical ideas on paper, followed by active parallel exploration on paper and in software, prior to final execution of their ideas as an original score. We conducted a participatory design study that focused on the creative exploration phase, to design tools that help composers better integrate their paper-based and electronic activities. We then developed InkSplorer as a technology probe that connects users? hand-written gestures on paper to Max/MSP and OpenMusic. Composers appropriated InkSplorer according to their preferred composition styles, emphasizing its ability to help them quickly explore musical ideas on paper as they interact with the computer. We conclude with recommendations for designing interactive paper tools that support the creative process, letting users explore musical ideas both on paper and electronically.

Wikibook: [22]. With the Wikibook project [22] we investigate how Wikibooks authors collaborate to create high-quality books. We combined Information Retrieval and statistical techniques to examine the complete multi-year lifecycle of over 50 high-quality Wikibooks. We found that: 1. The presence of redundant material is negatively correlated with collaboration mechanisms; 2. For most books, over 50% of the content is written by a small core of authors; and 3. Use of collaborative tools (predicted pages and talk pages) is significantly correlated with patterns of redundancy. Non-redundant books are well-planned from the beginning and require fewer talk pages to reach high-quality status. Initially redundant books begin with high redundancy, which drops as soon as authors use coordination tools to restructure the content. Suddenly redundant books display sudden bursts of redundancy that must be resolved, requiring significantly more discussion to reach high-quality status. These findings suggest that providing core authors with effective tools for visualizing and removing redundant material may increase writing speed and improve the book’s ultimate quality.

6.3. Engineering of interactive systems

Participants: Caroline Appert, Michel Beaudouin-Lafon, Anastasia Bezerianos, Olivier Chapuis, Jérémie Garcia, Stéphane Huot, Ilaria Liccardi, Wendy Mackay [correspondant], Emmanuel Pietriga.

We started working on jBricks [24], a Java toolkit enabling the exploratory prototyping of interaction techniques and rapid development of post-WIMP applications running on cluster-driven interactive visualization platforms such as the WILD wall display (Section 7.1). Research on cluster-driven wall displays has mostly focused on techniques for parallel rendering of complex 3D models. There has been comparatively little research effort dedicated to other types of graphics and to the software engineering issues that arise when prototyping novel interaction techniques or developing full-featured applications for such displays. To fill this gap, jBricks integrates a high-quality 2D graphics rendering engine and a versatile input configuration module into a coherent framework, hiding low-level details from the developer. The goal of this framework is to ease the development, testing and debugging of interactive visualization applications for wall-sized displays. It also offers an environment for the rapid prototyping of novel interaction techniques and their evaluation through controlled experiments, such as the one we recently conducted about mid-air pan-and-zoom techniques for wall-sized displays (see Section 6.1).

We developed the Shared Substance framework for multisurface interaction [20]. It is based on Substance, which implements a novel programming model called data orientation that separates functionality from data. Shared Substance extends Substance to distributed environments. It makes distribution explicit so that the programmer can dynamically add, reconfigure and remove components at runtime. An application built with Shared Substance is a collection of processes called environments that run on different machines. Environments are discovered dynamically and can appear and disappear at any time. Each environment contains a hierarchical data structure that can be shared, in whole or in part, with other environments. Sharing can be done through replication or mounting, which entail different performance trade-offs. Shared Substance also includes the Instrumental Interaction Kit (IIK) to facilitate the development of instruments in a multisurface environment [34]. [42]. We have used Shared Substance to develop several applications
for our WILD multisurface environment: *Substance Canvas* manages a virtual canvas that can span multiple interactive surfaces managed by different computers, such as the tiled display, interactive table and users’ laptops of the WILD room; Content can be added to the canvas from various an extensible set of sources, including live applications using Scotty (see below). *Substance Grise* wraps an existing application for displaying 3D brain scans into a Shared Substance environment; This allows us to run 64 copies of the application, each showing a different brain scan, and synchronize the 3D orientation of the scans using a brain prop that the user turns in his hand.

We explored the notion of user interface programming at run-time to create more malleable software [18]. Rather than creating a new user interface toolkit or supporting the customization of an interactive application from outside, we explored how well-defined hooks and a few high-level constructs could allow a programmer to modify an application “from inside”, i.e. using code that is dynamically loaded by the application at run-time. Compared with existing approaches, this supports deep customization that involve the behavior of the application, not just the surface of its user interface. The Scotty prototype implements run-time interface programming in the Mac OS X environment for any application written with the native Cocoa framework. We have used Scotty to distribute the user interface of an application over multiple devices, e.g. to move the Print button of an application onto an iPhone so the user can safely print while physically close to the printer; to replace a tool palette in an application with a toolglass; to check for the presence of attachments in an email application before sending an email; to add subtitles to a video viewer that does not have this functionality. We have also used Scotty in connection with the work on Shared Substance (see above) to teleport a live vector-based representation of a running applications to the WILD wall display. The advantage of this approach over, e.g., VNC, is that the content is properly scaled, taking advantage of the full resolution of the wall.
6. New Results

6.1. Efficient XML and RDF data management

Participants: Mohamed Amine Baazizi, Nicole Bidoit, Dario Colazzo, François Goasdoué, Konstantinos Karanasos, Asterios Katsifodimos, Julien Leblay, Noor Mallía, Ioana Manolescu, Alexandra Roatis, Marina Sahakyan, Federico Ulliana.

6.1.1. Materialized views for XML

We have continued our work on optimizing XML queries through materialized view-based rewriting, implemented within the ViP2P system. We published in IEEE ICDE 2011 an algorithm for rewriting XQuery queries using materialized XQuery view, which improves the state of the art in terms of expressive power of the supported XQuery subset, in collaboration with V. Vassalos (AUEB, Greece) \[41\]. Two follow-up works concern: efficient algebraic algorithms for incrementally maintaining the materialized views when the underlying documents change, in collaboration with A. Bonifati (CNR, Italy) \[35\], and algorithms for automatically recommending views to materialize for a given XML query workload, with V. Vassalos.

6.1.2. Type-based Update Optimization for XML

XML projection is a well-known optimization technique for reducing memory consumption for XQuery in-memory engines in order to overcome the main-memory limitations of these systems (Galax, Saxon, QizX, and eXist). One of our main research line focuses on a schema-based projection technique for for update optimization. The update language considered is XQuery Update Facility (XUF). The main idea behind this technique is: given a query $q$ over an XML document $t$, instead of evaluating $q$ over $t$, the query $q$ is evaluated on a smaller document $t'$ obtained from $t$ by pruning out, at loading time, parts of $t$ that are irrelevant for $q$. The queried document $t'$, a projection of the original one, is often much smaller than $t$ due to selectivity of queries.

The scenario and type-based projection proposed for XML queries, cannot be applied directly for updates. We have proposed a new scenario which is composed of four steps:

1. from the update $U$ and the DTD $D$, a type projector $P$ is inferred;
2. the document $t$, valid wrt $D$, is projected following $P$ in a streaming manner, at loading time;
3. $U$ is evaluated over the projection $P(t)$ and produces a partial result $U(P(t))$;
4. the initial document $t$ is merged with $U(P(t))$, in a streaming manner, at writing (serializing) time in order to produce the final result $U(t)$.

The scenario has been first studied and implemented for a kind of type projector which is a good compromise between simplicity and effectiveness, and corresponding results have been published in [31]. Subsequently, we have improved this technique by designing: (i) a new kind of type projector that minimizes the amount of data kept in the projection, and (ii) a new merge algorithm using the improved type projector. This analysis is complicated by the strong interconnection between the two tasks: while minimizing the projection we need to ensure a fast and correct merge process [17].

These results have also been presented in [34], providing an overview on the use of types and constraints from relational to XML data, and in the tutorial [63] focusing on schema-based techniques for safe and efficient XML processing.
6.1.3. XML query-update independence

A query and an update are independent when the query result is not affected by update execution, on any possible input database. Detecting query-update independence is of crucial importance in many contexts: view maintenance, concurrency, access control policies etc. Benefits are amplified when query-update independence can be checked statically. We propose a novel schema-based approach for detecting XML query-update independence. Differently from traditional schema-based analysis for XQuery, our system infers sequence of labels, called chains, that are vertically navigated in each schema instance by query and update paths. More precisely, for each node that can be selected by a query/update path in a schema instance, the system infers a chain recording: a) all labels that are encountered from the root to the selected node, and b) the order in which these labels are traversed. The contextual and ordering information provided these chains is at the basis of an extremely precise static independence analysis. We have devised a quite precise chain-inference system, and devised techniques for efficient implementation of the chain-based independence analysis. Results and experiments concerning this line of research have been recently submitted to an international conference.

6.1.4. Precision and complexity of XQuery type inference

A key feature of XQuery is its type system. Any language expression is statically typed and its type is used during program type-checking. In XQuery, types of input data and functions are defined in terms of regular expression types, but it is quite easy to write queries that generate non-regular languages. As a consequence, any type system for XQuery has to rely on a type inference process that approximates the (possibly non-regular) output type of a query with a regular type. This approximation process, while mandatory and unavoidable, may significantly decrease the precision of the inferred types. In [37] we study and compare in terms of precision and computational complexity two main existing XQuery type systems.

6.1.5. Managing temporal XML documents

The management of temporal data is a crucial issue in many database applications. We are currently investigating efficient storage and update methods for temporal XML documents, with a focus on compactness of the representation. One of the method developed relies on the type based optimization method developed for updates [31]. First results about this line of research are included in recent publications [29], [30]. Next research activities focus on the design of expressive temporal query and update languages, and on the use of techniques proposed in [29] for query and update optimization.

6.1.6. Materialized view selection for RDF

Syntactically, RDF, the data format of the Semantic Web, resembles relational data. However, RDF query processing is significantly complicated by the irregular nature of RDF data and by its simplistic data model, which leads to syntactically complex queries (involving many joins over the whole triple data set). When a query workload is known, the performance of the workload can be significantly improved by materializing access support data structures such as materialized views. Our efficient algorithms for selecting a set of views to materialize in order to speed up the processing of a set of RDF queries are described in a recent publication [21]. The prototype implementing them has been demonstrated at [53].

6.1.7. Hybrid models for XML and RDF

We have obtained interesting results in the area of jointly managing XML and RDF data. A first direction of work in this area was to support annotated documents, that is, XML documents where individual nodes or fragments could be annotated with RDF triples. This model allows to capture, for instance, blog comments, user ratings on social sites etc. We have proposed a general model based on W3C standards for modeling such data [39].

6.1.8. RDF query answering

The current trend for efficiently querying RDF datasets consists of delegating query evaluation to a scalable RDBMS. However, RDF query answering requires in addition to handle – outside the RDBMS – the RDF semantics. We have introduced the database (DB) fragment of RDF, encompassing the popular Description
Logic (DL) one with essential RDF features like modeling incomplete information, for which we have devised novel saturation- and reformulation-based techniques for answering the Basic Graph Pattern (BGP) queries of SPARQL. This extends the state of the art on pushing RDF query processing within RDBMSs.

6.2. Models for Web data management

Participants: Serge Abiteboul, Emilien Antoine, Meghyn Bienvenu, Alban Galland.

A book on Web Data Management and Distribution [54] was published this year.

6.2.1. A rule-based language for Web data management

We recently proposed [26] a Datalog-style rule-based language (called Webdamlog) for web data management. A novel feature of our language is delegation, that is, the possibility of installing a rule at another peer. In its simplest form, delegation is essentially a remote materialized view. In its general form, it allows peers to exchange rules, i.e., knowledge beyond simple facts, and thereby provides the means for a peer to delegate work to other peers.

A key contribution of our work is a study of the impact on expressiveness of delegations and explicit timestamps. We showed that both strictly augment the power of the language. In order to validate the semantics of our model, we demonstrated that under certain natural conditions, our semantics converges to the same semantics as the centralized system with the same rules.

6.2.2. Web information management with access control

We investigated the problem of sharing private information on the Web, where the information is hosted on different machines that may use different access control and distribution schemes. Based upon our work on Webdamlog, we introduced a distributed knowledge-base model, termed WebdamExchange, that comprises logical statements for specifying data, access control, distribution and knowledge about other peers. In a demo at ICDE [28], we showed how the model can be used in a social-network context to help users keep control on their data on the web. In particular, we demonstrated how users within very different schemes of data distribution (centralized, DHT, unstructured P2P, etc.) can still transparently collaborate while keeping a good control over their own data.

6.3. Ontology-based data and document Management

Participants: Meghyn Bienvenu, François Goasdoué, Yassine Mrabet, Nathalie Pernelle, Gianluca Quercini, Chantal Reynaud, Brigitte Safar, Fabian Suchanek.

6.3.1. Semantic Annotation

We have started a work on semantic annotation of public administration data in the setting of the project DataBridges, an ICT Labs activity. We considered public data represented in tables. The tables that we studied were tables created and published by INSEE. They are spreadsheets filled with statistics about geographic locations and are usually composed of multiple columns, of which one, that we term the subject column, contains a list of textual references to geographic entities, or toponyms, while the others contain numeric attributes. We proposed an approach and an algorithm that assigns a type, or header to the subject column of an INSEE table and identifies the geographic entities referred to by the toponyms in the column [64]. An external resource, DBpedia, is used to help to disambiguate the entities mentioned in the tables and a domain ontology ensures that the types are relative to the geographic domain. This work is continued in the setting of a post-doctoral work granted by the ANR project DataBridges. The aim of the project being to enrich a data warehouse, a first work is to automatically build an initial RDF data warehouse from data collected from the web.

6.3.2. Adaptive Ontologies for Information Retrieval

We published the approach supported by the TARGET framework for Web Information Retrieval in the International Journal of Web Portals (IJWP) [22]. This approach was the core of the PhD of Cédric Pruski defended in April 2009.
6.3.3. Querying ontology-based annotations

We have pursued our work on integrating knowledge bases and semantic annotations made on more or less structured tagged documents. We have defined an approach where RDF named graphs are used to distinguish uncertain semantic annotations from rdf triples that are provided by the populated ontology. A user domain query is then reformulated to obtain answers that are ranked according to their provenance (knowledge bases or annotations) [61].

6.3.4. Watermarking for ontologies

Ontologies are usually available under some type of license. The large ontologies of the Semantic Web, e.g., are commonly licensed under a Creative Commons License or a GNU license. These licenses require giving credit to the authors of the ontology if the ontology is ever used somewhere else. However, it can be hard to prove whether an ontology is used somewhere else, because ontologies contain world knowledge. If someone “steals” an ontology and uses it somewhere else, he can always claim that he collected the data by himself from real-world sources. To tackle this problem, we have studied approaches that watermark an ontology [43]. If a watermarked ontology is used somewhere else, the mark proves that the ontology has been stolen. Existing approaches have mainly modified the facts in the ontology to create a mark. This, however, compromises the precision of the ontology. Therefore, we have developed an approach that does not modify, but remove certain facts. Thereby, the precision of the ontology is not affected. We show that only a handful of facts have to be removed from an ontology to protect it against theft.

6.3.5. Consistent query answering in DL-Lite

An important problem which arises in ontology-based data access is how to handle inconsistencies. In the database community, the related problem of querying databases which violate integrity constraints has been extensively studied under the name of consistent query answering. The standard approach is based on the notion of a repair, which is a database which satisfies the integrity constraints and is as similar as possible to the original database. Consistent answers are defined as those answers which hold in all repairs. A similar strategy can be used for description logics by replacing the integrity constraints with the ontology. Unfortunately, recent work on consistent query answering in description logics has shown this problem to be co-NP-hard in data complexity, even for instance queries and the simplest DL-Lite dialect. In light of this negative result, we considered the problem of identifying cases where consistent query answering is feasible, and in particular, can be done using query rewriting, with the aim of better understanding the cases in which query rewriting can be profitably used. In [51], we make some first steps towards this goal by formulating general conditions which can be used to prove that a consistent rewriting does or does not exist for a given DL-Lite TBox and instance query.

6.3.6. Module-based data management in DL-lite

The current trend for building an ontology-based data management system (DMS) is to capitalize on efforts made to design a preexisting well-established DMS (a reference system). The method amounts to extract from the reference DMS a piece of schema relevant to the new application needs – a module –, possibly to personalize it with extra-constraints w.r.t. the application under construction, and then to manage a dataset using the resulting schema. We have revisited the reuse of a reference ontology-based DMS in order to build a new DMS with specific needs. We go one step further by not only considering the design of a module-based DMS (i.e., how to extract a module from a ontological schema): we also study how a module-based DMS can benefit from the reference DMS to enhance its own data management skills. We consider the setting of the DL-Lite\textsubscript{A} dialect of DL-Lite, which encompasses the foundations of the QL profile of OWL2 (i.e., DL-Lite\textsubscript{R}): the W3C recommandation for managing efficiently large datasets. We introduce and study novel properties of robustness for modules that provide means for checking easily that a robust module-based DMS evolves safely w.r.t. both the schema and the data of the reference DMS. From a module robust to consistency checking, for any data update in a corresponding module-based DMS, we show how to query the reference DMS for checking whether the local update does not bring any inconsistency with the data and the constraints of the reference DMS. From a module robust to query answering, for any query asked to a module-based
DMS, we show how to query the reference DMS for obtaining additional answers by also exploiting the data stored in the reference DMS.

### 6.4. Data and Knowledge Integration

**Participants:** Julio Cesar Dos Reis, Fayçal Hamdi, Rania Khefifi, Yassine Mrabet, Nathalie Pernelle, Chantal Reynaud, Fatiha Saïs, Brigitte Safar, Fabian Suchanek, Danai Symeonidou.

#### 6.4.1. Reference Reconciliation

The reference reconciliation problem consists in deciding whether different data descriptions refer to the same real world entity (same person, same conference etc.) Some of existing approaches, such as LN2R, are declarative and knowledge-based. Different kinds of knowledge can be declared in a domain ontology, like disjointness between classes or key constraints. This knowledge can be exploited to infer reconciliation and non-reconciliation decisions.

Our reference reconciliation work pursues three directions:

- **develop an automatic approach of key constraint discovery.** We have proposed in [46] KD2R, a method which allows automatic discovery of key constraints associated to OWL2 classes. These keys are discovered from RDF data which can be incomplete. The proposed algorithm allows this discovery without having to scan all the data. KD2R has been tested on data sets of the international contest OAEI and obtains promising results.

- **develop a reference reconciliation method for detecting redundant data in case of web data tables that are semantically annotated by an ontology.** Each table cell values consists in numerical fuzzy set (NFS) or in symbolic fuzzy set (SFS). We have developed a method which uses ontology knowledge and computes similarity scores to decide the data redundancy. We have also proposed two similarity measures for numerical fuzzy set as well as symbolic fuzzy set. The proposed measures are more flexible than existing ones. This approach has been published in [36], [58]. We are working on its extension to be able to distinguish redundant data from similar ones by using provenance information.

- **develop a new approach which addresses the problem of resource discovery in the Linked Open Data cloud (LOD) where data described by different schemas is not always linked.** We have proposed an automatic approach in [42], [58] that allows discovery of new links between data. These links can help to match schemas that are conceptually relevant with respect to a given application domain. Furthermore, these links can be exploited during the querying process in order to combine data coming from different sources. In this approach we exploit the semantic knowledge declared in different schemas in order to model: (i) the influences between concept similarities, (ii) the influences between data similarities, and (iii) the influences between data and concept similarities. The similarity scores are computed by an iterative resolution of two non linear equation systems that express the concept similarity computation and the data similarity computation.

#### 6.4.2. Context-aware Personal Information Management

Personal information management (PIM) is the practice and analysis of the activities performed by people to acquire, organize, maintain, and retrieve information for everyday use. PIM is a growing area of interest because, everyone is looking for better use of our limited personal resources of time, money and energy. Several research on the topic is being done in different disciplines, including human-computer interaction, database management, information retrieval and artificial intelligence.

The increasingly big amount of personal information (e.g., mails, contacts, appointments) managed by a user is characterized by their heterogeneity, their dispersion and their redundancy. The general goal of this work consists in designing a system, which allows providing the end-users personal data access with services that are relevant to his/her needs, and to access personal data both by mobile devices (smartphone) and Internet-connected Personal Computers. More specifically, we focus here on the problem of defining a common
meta-model for a flexible and homogeneous personal information management. The meta-model that we propose allows users creating personal information and organizing them according to different points of view (ontologies) and different contexts. Contextual queries are defined to allow users to retrieve its personal information using the geographical contexts. The semantic Web languages (OWL, RDF and SPARQL) are used to implement the approach.

6.4.3. Mapping between ontologies

We pursue our work on ontology alignment in the setting of the ANR GeOnto project by aiming to provide full life-cycle support for ontologies.

We investigated how alignment results generated by our alignment tool, TaxoMap, can be used to enrich one ontology with another. We shown that the enrichment process depends on characteristics of the ontology used for enrichment. Three enrichment contexts identified in the setting of the ANR project GeOnto have been studied and enrichment treatments performed. A first context considers ontologies of the same application domain and of a reasonable size. A second context considers small ontologies previously extracted from a generalist one. A third context considers enrichment from a huge, generalist ontology, such as Yago. Early results obtained in the setting of the ANR project GeOnto in the topographic domain have been published in [50], [25].

The module supporting our enrichment approach has been implemented in TaxoMap Framework using patterns. Initially, TaxoMap Framework was composed of our alignment tool, TaxoMap, we are working on for several years in the team and of a mapping refinement module. We extended it in order to obtain a broader framework and an interactive environment by including TaxoPart, a partitioning tool we developed to split two huge ontologies which could not be aligned into two sets of blocks of a limited size, and a module specific to ontology enrichment. Moreover, we re-implemented TaxoMap, our alignment tool, as a web service to make it easily accessible at: http://taxomap.lri.fr:8000/axis2/services/TaxoMapService?wsdl.

We also started a PhD work, joined with CRP Henri Tudor in Luxembourg, to investigate issues dealing with medical knowledge organizing systems evolution. We will define a formal framework to support medical knowledge organizing systems evolution in a consistent way and also to support the maintenance of mappings directly impacted by knowledge organizing systems local evolution.

On a related topic, we have developed a probabilistic framework, PARIS (Probabilistic Alignment of Relations, Instances and Schema), for matching ontologies holistically, thereby exploiting synergies between matches on the instance level and matches on the schema level [57]. The framework is parameter-free and does not require resource-specific tuning. PARIS is fully implemented and has been shown to match some of the largest ontologies on the Semantic Web with a precision of around 90%.

6.4.4. Integration of Web resources

We have pursued our work on integration of resources available on the Web in Adaptive Hypermedia Systems (AHS), allowing creators to define their own adaptation strategies based on their own domain models. The approach is based on a set of 22 adaptation patterns, independent of any application domain and independent of any adaptation engine, published in [59], [47]. These elementary adaptation patterns are organized in a typology in order to facilitate their understanding and their use in the EAP framework to define complex strategies. In [24], we described the whole process to generate complex adaptation strategies and how the generated strategies can be integrated into existing AHSs. The results of an experiment conducted in the e-learning domain is presented. It showed that the pattern-based approach for defining adaptation strategies is more suitable than those based on “traditional” AH languages.

We also pursued our work on the integration of the EAP framework and other AHSs. Our collaboration with A. Cristea from the University of Warwick (UK) led us to a very detailed study of adaptation languages. The first flexible generic adaptation language is the LAG adaptation language. We studied the expressivity of this initial adaptation language in comparison with our newly proposed language, in the EAP framework, and the pros and cons of various decisions in terms of the ideal way of defining an adaptation language. We proposed a unified vision of adaptation and adaptation language. The unified vision is not limited to the
two languages analyzed, and can be used to compare and extend other approaches in the future. Beside this theoretical qualitative study, we also made experimental evaluation and comparison of the two languages, and an article is currently being evaluated.

We have also investigated integration of Web services. The Search Computing project (“SeCo”) at the Polytechnic University of Milan aims to orchestrate Web services to answer user queries. Currently, the project represents Web services by so-called Service Marts. These are frame-like representations of the services, which follow the slot-value paradigm. This representation faces several challenges if more Web services get added to the system, because it is hard to ensure that Web services added by different users can still be joined. Therefore, we have explored a more ontological representation of Web services. In our proposal [ 55 ], Web services are represented as sub-graphs of an ontology. This allows users to add new Web services that re-use the vocabulary of existing Web services.

On related topics, together with researchers from the Max-Planck Institute in Saarbrucken, we have worked on extending the YAGO ontology. YAGO already contains dozens of millions of facts. With the present work, we aim to give these facts a temporal and a spatial dimension. For every event and every entity, we want to know where and when these objects existed. For this purpose, we have developed a methodology that extracts these types of facts from Wikipedia. We have also developed a logical reasoning framework that allows propagating these time and space annotations from some facts to others. This has grown YAGO to 80 million facts in total, making it an ontology that is anchored in time and space (Best demo award at the WWW 2011 conference [ 40 ]).

6.5. Reasoning over Distributed Systems


6.5.1. Distributed Diagnosis Problems

We pursued the work on distributed algorithms for diagnosing distributed systems. The general framework is consistency-based diagnosis for propositional-logic theories in a P2P setting with privacy constraints. It boils down to distributed implicant finding and is thus in some sense dual to the problem of consequence finding described in next paragraph. Vincent Armant is finishing is PhD and has extended his previous work on more general topics, i.e. focusing on the construction of a good decomposition of the network that will ensure an efficient reasoning mechanism. An important effort has been put in the design of a real-world sized experimentation on distributed systems.

Lina Ye defended her PhD on diagnosability analysis of distributed discrete-event systems, modeled as synchronized labeled automata. The aim of diagnosability is to ensure that a given partially observable system has the property that any fault (taken from a set of faults given a priori) will be detectable and identifiable without ambiguity in a finite time after its occurrence. Distributed diagnosability analysis is optimized by abstracting necessary and sufficient information from local objects to achieve global decision. After having addressed the distribution of the system’s model into local models, we focus in 2011 on the extension to systems where the observable information itself is distributed instead of centralized. Joint diagnosability definition has been provided and undecidability of deciding it has been proved in the general case where communication events are not observable, before proposing an algorithm to test its sufficient condition. In addition, decidability result and algorithm have been given when communications are observable.

Michel Batteux defended also his PhD (led in the framework of a CIFRE thesis with Sherpa Engineering) about diagnosability and diagnosis of technological systems. The work was led in the centralized case, focusing on defining, implementing, testing and validating on a real case study (a fuel cell system) an all-in-one tool to design a diagnosis system for technological systems by integrating representation of the system and its potential faults, off-line diagnosability analysis and automatic generation of the on-line embedded diagnoser.
6.5.2. Distributed Consequence Finding

A major reengineering of the SOMEWHERE platform, for decentralized consequence finding, has been initiated within the DISQUE project. Current efforts have focused on the rewriting of the communication layer, that now relies on the JXTA middleware. A new tool is also being developed, in order to facilitate large scale experimentations on a grid (Grid5000). This tool is designed in a fairly generic way, in order to be reusable for similar projects that require deploying sets of collaborating reasoners in a decentralized setting, and automating collaborative problem solving on various instances. We also expect this tool to be used for automating integration tests during further developments.

6.5.3. Towards distributed architectures for Modern SAT Solvers

If we aim at proposing a new architecture for distributed SAT Solvers, we pursued this year the improvements of Glucose, our centralized SAT solver. Glucose 2 won 3 medals at the SAT 2011 Competition, and one in the category Application SAT+UNSAT. We target to make a massively distributed version of Glucose, for very hard SAT problems.
6. New Results

6.1. Théorie spectrale max-plus et géométrie métrique/Max-plus spectral theory and metric geometry

6.1.1. Introduction

Participants: Marianne Akian, Stéphane Gaubert, Cormac Walsh.

Étant donné un noyau \( a : S \times S \to \mathbb{R} \cup \{-\infty\} \), on peut lui associer le problème spectral max-plus

\[
\sup_{y \in S} a(x,y) + u(y) = \lambda + u(x), \quad \forall x \in S,
\]

(1)
dans lequel on cherche le vecteur propre \( u : S \to \mathbb{R} \cup \{-\infty\} \) et la valeur propre correspondante \( \lambda \in \mathbb{R} \cup \{-\infty\} \). Comme nous l’avons rappelé dans les § 3.2 et 3.3, le problème spectral (9) intervient en contrôle ergodique: l’ensemble \( S \) est l’espace des états, et l’application \( a(x,y) \) fournit le gain associé à la transition \( x \to y \). Le cas où \( S \) est fini est classique, l’on a alors un résultat précis de représentation de l’espace propre, à l’aide d’un certain graphe, dit graphe critique. Des résultats existent également lorsque \( S \) est compact et que le noyau vérifie certaines propriétés de régularité.

Dans [79], nous avons considéré le cas où \( S \) est non compact. Lorsque \( \lambda = 0 \), l’espace propre est analogue à l’espace des fonctions harmoniques défini en théorie (classique ou probabiliste) du potentiel. En introduisant l’analogue max-plus de la frontière de Martin, nous avons obtenu un analogue de la formule de représentation de Poisson des fonctions harmoniques : toute solution \( u \) de (9) peut être représentée sous la forme :

\[
u = \sup_{w \in \mathcal{M}_m} w + \mu_u(w),
\]

(2)
où \( \mathcal{M}_m \subset (\mathbb{R} \cup \{-\infty\})^S \) est l’analogue max-plus de la frontière de Martin minimale (l’ensemble des fonctions harmoniques extrémales normalisées), et où \( \mu_u \) joue le rôle de la mesure spectrale. Nous avons montré aussi que les éléments de l’espace de Martin minimal peuvent être caractérisés comme les limites de “quasi-géodésiques”. La frontière de Martin max-plus généralise dans une certaine mesure la frontière d’un espace métrique construite à partir des horo-fonctions (fonctions de Busemann généralisées), ou horo-frontière. Ces résultats inspirent les travaux des sections suivantes, qui portent sur des cas remarquables d’espaces métriques (§ 6.1.4) ou sur des applications en théorie des jeux (§ 6.1.2).

English version

Let the kernel \( a : S \times S \to \mathbb{R} \cup \{-\infty\} \) be given. One may associate the max-plus spectral equation (9), where the eigenvector \( u : S \to \mathbb{R} \cup \{-\infty\} \) and the eigenvalue \( \lambda \in \mathbb{R} \cup \{-\infty\} \) are unknown. As we recalled in § 3.2 and refinemonotone, this spectral problem arises in ergodic optimal control: the set \( S \) is the state space, and the map \( a(x,y) \) is the transition reward. The case when \( S \) is finite is classical, a precise spectral theorem is known, with a characterization of the eigenspace in terms of a critical graph. Some results have been shown when \( S \) is compact, assuming that the kernel \( a \) satisfies some regularity properties.
In [79], we considered the case where \( S \) is non-compact. When \( \lambda = 0 \), the eigenspace is analogous to the set of harmonic functions defined in classical or probabilistic potential theory. By introducing a max-plus analogue of the classical Martin boundary, we obtained an analogous to the set of normalised extremal harmonic functions, and \( \mu_u \) plays the role of the spectral measure. We also showed that the elements of the minimal Martin boundary can be characterised as limits of certain “almost-geodesics”. The max-plus Martin boundary generalises to some extent the boundary of metric spaces defined in terms of horofunctions (generalised Busemann functions), or horoboundary. These results have inspired the work of the next sections, which deal either with remarkable examples of metric spaces (§ 6.1.4) or applications to zero-sum games (§ 6.1.2).

6.1.2. Une caractérisation maximin du taux de fuite d’applications nonexpansives/A maximin characterization of the escape rate of nonexpansive mappings

Participants: Stéphane Gaubert, Guillaume Vigeral.

Le problème de l’existence du gain moyen par unité de temps pour des jeux répétés à somme nulle conduit à étudier la limite \( f^k(v)/k \) lorsque l’horizon \( k \) tend vers l’infini, où \( f \) (l’opérateur de programmation dynamique) est une application contractante au sens large sur un espace de Banach, voir § 3.3. La limite peut ne pas exister, mais un résultat de Kohlberg et Neyman montre qu’il existe toujours une forme linéaire \( \phi \) de norme 1 telle que la limite de \( \phi(f^k(x)/k) \) existe lorsque \( k \to \infty \) et coïncide avec le “taux de fuite” \( \lim_{k \to \infty} \|f^k(x)/k\| \). Dans [29], nous avons généralisé ce résultat au cas où \( f \) est une application nonexpansive (contractante au sens large) sur un espace métrique vérifiant une forme affaiblie de l’hypothèse de courbure nonpositive au sens de Busemann. La forme linéaire \( \phi \) est alors remplacée par une horofonction, et l’on obtient une caractérisation de type “maximin” du taux de fuite, qui étend la formule de Collatz-Wielandt en théorie de Perron-Frobenius (le vecteur propre apparaissant dans cette formule correspond à une famille d’horoboules invariantes par la dynamique). Ceci est motivé par les problèmes de contrôle ou de jeux quadratiques, dans lesquels l’espace métrique est le cône des matrices définies positives muni de sa métrique Riemannienne invariante ou de la métrique de Thompson.

English version

The problem of the existence of the mean payoff per time unit for repeated games leads to studying the existence of the limit of \( f^k(v)/k \), where \( f \) is a nonexpansive map (the dynamic programming operator) acting on a Banach space, see § 3.3 for more background. The limit may not exist, but a result of Kohlberg et Neyman shows that there is always a norm one linear form \( \phi \) such that the limit of \( \phi(f^k(x)/k) \) exists and coincides with the limit of \( \|f^k(x)/k\| \) as \( k \) tends to infinity (the escape rate). In [29], we extend this result to the case of a nonexpansive map defined on a metric space satisfying a mild form of Busemann nonpositive curvature condition. Then, the linear form \( \phi \) is replaced by an horofonction, and we obtain a maximin type characterization of the escape rate, which extends the Collatz-Wielandt formula in Perron-Frobenius theory. This is motivated by the study of quadratic optimal control and game problems, in which the metric space is the cone of positive semi-definite matrices equipped with the Riemannian invariant metric or with Thompson metric.

6.1.3. Isométries de la géométrie de Hilbert/Isometries of the Hilbert geometry

Participants: Cormac Walsh, Bas Lemmens [Kent University, UK].

L’un des intérêts de l’horofrontière est de renseigner sur le groupe des isométries d’un espace métrique. En effet, ce groupe agit naturellement sur l’horofrontière, et cette action peut parfois être mieux comprise que l’action du groupe sur l’espace d’origine.

Nous étudions le groupe des isométries pour la métrique de Hilbert. De La Harpe [188] a donné plusieurs conjectures relatives à ce groupe. Nous conjecturons que le groupe des isométries est exactement le groupe des transformations linéaires projectives à moins que le domaine ne soit une coupe d’un cône symétrique non-Lorentzien. Nous avons démontré cette conjecture lorsque le domaine est un polytope [32].
Dans le cas général, on prouve, en utilisant les horo-fonctions, que si il existe une bijection entre deux cônes homogène de degré $-1$, antitone, et d’inverse antitone, ces deux cônes sont symétriques. Nous essayons maintenant de montrer que toute isométrie de Hilbert sur un domaine convexe est la version projective d’un automorphisme linéaire du cône sur le domaine, ou d’une bijection du cône, homogène de degré $-1$, qui est antitone et d’inverse antitone. Ce résultat permettrait de completer la preuve de la conjecture proposée plus haut.

**English version**

One use for the horofunction boundary is to study the group of isometries of a metric space. This is because this group has a well defined action on the horoboundary and it is likely that in many cases this action will be easier to understand than the action on the space itself. We have been investigating the isometries of the Hilbert geometry. De La Harpe [188] has previously made several conjectures about the isometry group of this space. We conjecture that the isometry group is exactly the group of projective linear transformations unless the domain on which the geometry is defined is a cross section of a non-Lorentzian symmetric cone. We have previously proved that this conjecture is true in the case of a polytope domain [32].

In the general case, we can now prove, using horofunctions, that if a bijection between cones is homogeneous of degree $-1$, order inverting, and has an order inverting inverse, then both cones are symmetric. We are working on showing that every Hilbert isometry on a convex domain arises by considering projectively either a linear automorphism on the cone over the domain, or a homogeneous $-1$, order inverting bijection on this cone with order inverting inverse. Establishing this result would complete our proof of the above conjecture.

**6.1.4. Espace de Teichmüller/Teichmüller space**

**Participant:** Cormac Walsh.

L’espace de Teichmüller d’une surface est un espace métrique composé des structures conformes de cette surface. On peut le voir comme l’ensemble des classes d’équivalence des métriques riemanniennes de cette surface, où deux métriques sont équivalentes si il existe une application conforme homotope à l’identité qui envoie l’une des métriques sur l’autre.

Il existe plusieurs métriques naturelles sur l’espace de Teichmüller. Nous avons travaillé précédemment sur la métrique Lipschitz de Thurston et avons prouvé [70] que l’horofrontière de cet espace métrique était la frontière de Thurston.

Néanmoins, la métrique la plus utilisées sur l’espace de Teichmüller est la métrique de Teichmüller. L’horofrontière de cet espace métrique n’est autre que la frontière déjà introduite dans la littérature sous le nom de frontière de Gardiner–Masur. Nous étudions cette frontière, en particulier nous donnons explicitement ses points de Busemann.

Par la suite, nous avons l’intention d’utiliser cette propriété afin d’étudier les sous-groupes du groupe modulaire, qui est le groupe des isométries de la métrique de Teichmüller.

**English version**

An interesting metric space is the Teichmüller space of a surface. This is the space of conformal structures on the surface. One may think of it as the space of equivalence classes of Riemannian metrics on the surface, where two such metrics are regarded as being equivalent if there is a conformal map on the surface taking one to the other that is homotopic to the identity.

There are several natural metrics on Teichmüller space. Previously, we have worked with Thurston’s stretch metric and have shown [70] that the horofunction boundary with this metric is just the usual Thurston boundary.
However, the most commonly used metric on Teichmüller space is Teichmüller’s metric. The horofunction boundary of this metric space turns out to be the same as a previously defined boundary, called the Gardiner–Masur boundary. We have been investigating this boundary, in particular we have managed to work out explicitly its Busemann points.

In future work, we intend to apply this knowledge to study subgroups of the mapping class group, which is the isometry group of the Teichmüller metric.

6.2. Algèbre linéaire max-plus et convexité abstraite/Max-plus linear algebra and abstract convex analysis

6.2.1. Convexité max-plus ou tropicale/Max-plus or tropical convexity

Participants: Xavier Allamigeon, Stéphane Gaubert, Eric Goubault [CEA], Ricardo Katz [Conicet, Argentine].

On étudie les analogues max-plus ou tropicaux des ensembles convexes. Ceux-ci sont utiles en particulier pour représenter de manière effective les ensembles d’états accessibles de systèmes à événements discrets [9], ils sont aussi apparus récemment en géométrie tropicale, dans toute une série de travaux à la suite de Sturmfels et Develin [114]. Les polyèdres max-plus peuvent aussi être vus comme des limites de déformations de polyèdres classiques, sur lesquels ils donnent un éclairage de nature combinatoire. Toutes ces motivations ont inspiré la recherche d’analogues des résultats fondamentaux d’analyse convexe classique: séparation, projection, points extrémaux, à la suite en particulier de [8].


On en déduit un analogue tropical de la méthode de la double description [57] (méthode très utilisée sur les polyèdres classiques, et dûe à Motzkin et al. [158]). Cet algorithme permet de calculer les sommets d’un polyèdre défini de façon externe (intersection de demi-espaces ou d’hyperplans tropicaux). Grâce au critère combinatoire précédent, l’algorithme améliore de plusieurs ordres de grandeur les techniques connues jusqu’alors. Ceci est confirmé par de nombreuses expérimentations. Ce travail est motivé par des applications à l’analyse statique [82] et aux systèmes à événements discrets [116], dans lesquelles la manipulation de tels polyèdres est le goulot d’étranglement.

Dans un travail de X. Allamigeon, S. Gaubert, et R. Katz [57], on étend le théorème de McMullen au cas tropical: ce dernier caractérise le nombre maximal de points extrémaux d’un polyèdre, en fonction du nombre d’inégalités qui le définissent et de sa dimension. Nous montrons que la même borne est valide dans le cas tropical (à une modification triviale près). Cependant, le calcul de la borne optimale est encore ouvert dans ce cas.

Dans un travail de S. Gaubert et R. Katz [26], on étudie la représentation d’un polyèdre tropical comme intersection de demi-espaces, ou si l’on préfère, comme conjonction d’inégalités affines. Nous donnons notamment un contre-exemple, montrant les inconvénients de la représentation en termes de demi-espaces minimaux proposée précédemment dans la littérature tropicale.

We study the max-plus or tropical analogues of convex sets. These have been used in particular to represent effectively the accessible sets of certain discrete event systems [9]. They also appeared in tropical geometry, following the work of Sturmfels and Develin [114]. Max-plus polyhedra can be thought of as limits of deformations of classical polyhedra, on which they give a combinatorial insight. These motivations have inspired the investigation of analogues of basic results of classical convex analysis: separation, projection, representation by extreme points, following [8].

In a work of X. Allamigeon, S. Gaubert, and E. Goubault [57], we introduce a combinatorial criterion for the characterization of the vertices of tropically convex polyhedra. It is expressed in terms of directed hypergraphs and their strongly connected components. This criterion can be verified in almost linear time in the size of the hypergraph.

This allows to develop a tropical analogue of the double description method [57] (this method is widely used for classical convex polyhedra, and is due to Motzkin et al. [158]). This algorithm is able to determine all the vertices of a polyhedron defined externally (intersection of tropical half-spaces of hyperplanes). Thanks to the combinatorial criterion mentioned above, the algorithm improves the existing methods by several orders of magnitude. This is confirmed by several experiments. This is motivated by applications to static analysis [82] and discrete event systems [116], in which computing such polyhedra turns out to be the bottleneck.

In a work of X. Allamigeon, S. Gaubert, and R. Katz [19], we extend the McMullen upper bound theorem to the tropical case. This theorem characterizes the maximal number of extreme points of a polyhedron, as a function of the number of inequalities defining it, and of the dimension. We show that the same bound is valid in the tropical case (up to a trivial modification). However, computing the optimal bound is an open problem in this case.

In a work of S. Gaubert and R. Katz [26], we study the representation of a tropical polyhedron as an intersection of half-spaces. We give in particular a counter example, showing some inconvenients of the representation in terms of minimal half-spaces proposed previously in the tropical literature.

It is well-known that a tropical polyhedron can be represented as the convex hull of a minimal set of points and rays, provided by its vertices and extreme rays [125]. In an ongoing work of X. Allamigeon and R. Katz, partly done during the visit of R. Katz at INRIA (July 2011), the dual problem of characterizing the minimal representations by half-spaces is studied. We show that a tropical polyhedron admits essentially a unique minimal external representation by half-spaces, provided that their apices belong to the polyhedron. We prove that the apices of these half-spaces correspond to certain vertices of the tropical complex introduced by Develin and Sturmfels [114]. We also establish a combinatorial criterion allowing to eliminate redundant half-spaces using directed hypergraphs.

### 6.2.2. Convexes max-plus et jeux avec paiements ergodiques/Max-plus convex sets and mean payoff games

**Participants:** Marianne Akian, Xavier Allamigeon, Stéphane Gaubert, Alexander Guterman [Moscow State University], Ricardo Katz [Conicet, Argentine], Sergei Sergeev.

Dans un travail d’Akian, Gaubert et Guterman [16], on montre un résultat d’équivalence entre les jeux ergodiques à somme nulle et les systèmes d’inégalités max-plus linéaires: décider la non-vacuité d’un polyèdre tropical est équivalent à vérifier si un jeu déterministe à somme nulle a un paiement moyen par unité de temps positif ou nul. Plus généralement, la même question pour un jeu stochastique à somme nulle est équivalente à vérifier si un convexe tropical (non-polyédral, i.e., défini par un système infini d’inégalités) est vide. Ces résultats sont démontrés à l’aide de techniques de théorie de Perron-Frobenius non-linéaire. Ils sont ensuite appliqués à l’étude de l’indépendance linéaire dans le semi-anneau tropical.

Le résultat de [16] a eu plusieurs retombées.
D’une part, dans un travail d’Allamigeon, Gaubert, et Katz [20], on établit un analogue tropical du théorème de Farkas: on montre que décider si une inégalité max-plus linéaire est une conséquence logique d’une famille de telles inégalités est également équivalent à un problème de jeu ergodique. Le travail [20] comprend aussi une description des “faces” (ou plus précisément, des points extrêmes du polaire) d’un polyèdre tropical en termes de transversaux minimaux dans un hypergraphe.

D’autre part, dans un travail de Gaubert et Sergeev [127], on réduit le problème spectral tropical de type faisceaux, $Ax = \lambda Bx$, à un jeu paramétrique (ce qui permet de calculer le spectre en temps pseudo-polynomial).

Enfin, dans un travail de Gaubert, Katz, et Sergeev [27], on développe un algorithme de programmation linéaire tropicale (pseudo-polynomial) basé sur cette correspondance avec les jeux répétés. Allamigeon et Sergeev ont développé récemment un logiciel prototype en ocaml afin d’expérimenter cet algorithme. Ce logiciel inclut en particulier, en guise d’oracles résolvant des problèmes de jeux répétés, l’algorithme d’itération sur les politiques de Gaubert-Gunawardena [124] dans sa version de [115], la modification de cet algorithme due à Chaloupka [98], ainsi que l’algorithme d’itérations sur les politiques de Björklund-Vorobyov [91].

### English version

In [16], we show the equivalence mean payoff games and max-plus linear inequalities: testing whether a tropical polyhedron is non-empty is equivalent to checking whether a mean payoff deterministic game is winning. More generally, checking whether a mean payoff stochastic game is winning is equivalent to checking the non-emptiness of a tropical convex set defined by an infinite family of inequalities. These results are established using techniques of non-linear Perron-Frobenius theory. Then, they are applied to the study of linear independence over the tropical semiring.

The equivalence established in [16] had several consequences.

First, a work of Allamigeon, Gaubert, and Katz [20] yields a tropical analogue of Farkas theorem: we show that deciding whether a max-plus linear inequality follows from a family of such inequalities is also equivalent to solving a mean payoff game. Moreover, the work [20] comprises a characterization of the “faces” (more precisely, the extreme points of the polar) of a tropical polyhedra in terms of minimal transversals of a hypergraph.

Next, in a work of Gaubert and Sergeev [127], the tropical spectral problem for matrix pencils, $Ax = \lambda Bx$, is reduced to a parametric game (which allows one to compute the spectrum in pseudo-polynomial time).

Finally, in a work of Gaubert, Katz, and Sergeev [27], a (pseudo-polynomial) tropical linear programming algorithm is developed, based on the same correspondence with mean payoff games. Allamigeon and Sergeev developed recently an ocaml prototype in order to experiment this method. This prototype includes in particular mean payoff game solvers, namely the version of [115] of the policy iteration algorithm of Gaubert-Gunawarden [124], the modification of this algorithm by Chaloupka [98], as well as the policy iteration algorithm of Björklund-Vorobyov [91].

### 6.2.3. Meilleure approximation par des semi-modules max-plus pour la métrique projective de Hilbert/Best approximation in Hilbert’s projective metric by max-plus semimodules

**Participants:** Marianne Akian, Stéphane Gaubert, Viorel Nitica [West Chester University (US) and IMAR (Bucharest, Romania)], Ivan Singer [IMAR (Bucharest, Romania)].

Nous étudions les projecteurs sur des espaces linéaires max-plus, ainsi que les demi-espaces max-plus séparants. Dans [18], nous obtenons de nouvelles propriétés concernant ces objets, ce qui nous permet de déduire une formule explicite de la distance d’un point à un demi-espace max-plus pour la métrique projective de Hilbert, ainsi qu’une caractérisation de l’ensemble des points minimisants de cette distance. Nous obtenons aussi un algorithme de type projection cyclique permettant de résoudre des systèmes d’inégalités linéaires max-plus. Ce travail est effectué dans le cadre d’un projet LEA Math mode.
We are studying projectors on max-plus linear spaces, as well as separating half-spaces over the max-plus semiring. In [18], we establish new results, and derive an explicit formula for the distance in Hilbert’s projective metric between a point and a half-space, as well as explicit descriptions of the set of minimizers of this distance. We also obtain, as a consequence of the previous results, a cyclic projection type algorithm to solve systems of max-plus linear inequalities. This work is carried out as part of a LEA Math-mode project.

6.2.4. Miscellanées en algèbre linéaire max-plus/Topics in max-plus linear algebra

Participant: Sergei Sergeev.

During his one year post-doctoral stay, S. Sergeev collaborated with team members on tropical convexity and mean payoff game problems (§ 6.2.2 supra). In addition, the following research has been developed with several international coauthors.

- **Z-matrix equations and weakly stable matrices.** Sergeev continued his joint research with P. Butkovič and H. Schneider on traditional topics of max-plus linear algebra. They described solution set to Z-matrix equations $\lambda x = Ax \oplus b$, comparing the results with nonnegative matrix setting (old works of H. Schneider with coauthors) and extending them to linear algebra over more general semirings. **Weakly stable matrices** are such that the set of vectors $x$ whose orbit $A^k x$ converges to an eigenvector, is exactly the eigenvector cone. The weak stability was characterized in terms of relations between spectral classes of $A$ and the critical graph in each strongly connected component.

- **CSR expansions.** Sergeev continued his research on CSR expansions of matrix powers in max-plus algebra, see the accepted paper with H. Schneider [37], and paper with T. Nowak in preparation, which is going to be a survey on transience bounds in max-plus algebra based on CSR expansion schemes. New results may be applied in analysing the performance of reversal routing and reversal scheduling algorithms in collaboration with the team of B. Charron-Bost (École Polytechnique).

- **Ultradiscrete KdV.** Sergeev learned about the ultradiscrete KdV model at the conferences in Manchester (April 2011) and Glasgow (July 2011), and worked on the application of max-plus spectral theory to the ultradiscrete analogue of the Lax pair studied recently by R. Willox, J. Satsuma, J. Nimmo and others, based on the idea of S. Gaubert (who also took part in both conferences). In submission [66], he suggested the notion of pairs of fundamental eigenvectors associated with each soliton and showed that the problem can be reduced to finite-dimensional spectral theory with two additional constraints. In some special cases the problem can be solved by means of fundamental eigenvectors, which can be also used in the operation of undressing. However, in the case of several massive solitons the fundamental eigenvectors can never provide a solution, and more elaborate theory is needed.

- **max-Łukasiewicz algebra** Sergeev learned about some new fuzzy linear algebras during his research visit to Czech Republic. He observed that the linear algebraic problems over max-Łukasiewicz algebra, defined in the interval $[0, 1]$ with operations $a \otimes b := \max(a + b - 1, 0)$ and $a \oplus b := \max(a, b)$ can be reduced to some problems over max-plus algebra. This observation, being used in submission [63] has led to a new collaboration with Martin Gavalec and his group at the University of Hradec Kralove.

- **Visualization scaling and multi-objective optimization.** Sergeev continued his research on diagonal similarity scalings in max-plus algebra, see submission [67] and new collaboration with B. Benek Gursoy and O. Mason (in preparation). In this new collaboration, the authors analyze the possibilities of simultaneous matrix scaling (visualization), the special case of symmetrically reciprocal matrices, and extension to Pareto optimality.
6.3. Algèbre max-plus, déformations et asymptotiques /Max-plus algebra, deformations and asymptotic analysis

6.3.1. Introduction

Comme indiqué dans le § 3.7, l’algèbre max-plus est la limite d’une déformation de l’algèbre classique, ou plutôt du semi-corps des réels positifs. Elle peut aussi fournir des estimations de ces déformations, puisque

\[ \max(a, b) \leq \epsilon \log \left( e^{a/\epsilon} + e^{b/\epsilon} \right) \leq \epsilon \log (2) + \max(a, b). \]

(3)

L’utilisation de ces propriétés a déjà conduit dans le passé aux travaux sur les perturbations de valeurs propres [75], [74], [73], ou sur les grandes déviations [1], [77]. Dans les travaux qui suivent, nous exploitons ces propriétés dans des contextes reliés ou similaires à ceux de nos travaux précédents.

\textbf{English version}

As detailed in § 3.7, max-plus algebra is the limit of a deformation of classical algebra, or more precisely of the semi-field of usual real positive numbers. It can also give estimations for these deformations using for instance (11). By using these properties, we already obtained some works on singular perturbations of matrix eigenvalues [75], [74], [73], or on large deviations [1], [77]. In the works described below, we are exploiting again these properties in contexts that are related or similar to those of our earlier works.

6.3.2. Aspects tropicaux des algorithmes de scaling matriciel/Tropical aspects of matrix scaling problems

Participants: Marianne Akian, Stéphane Gaubert, Laura Grigori, Meisam Sharify Najafabadi.

Le travail de thèse de M. Sharify [13] a porté sur les méthodes de mise à l’échelle utilisées en algorithmique numérique matricielle pour améliorer la précision des calculs.

Une première partie du travail, appliquant les techniques de [73], [74], porte sur les problèmes de valeurs propres. On montre notamment que l’ordre de grandeur des valeurs propres d’un faisceau matriciel est donné (sous des conditions de non-dégénérescence) par les valeurs tropicales, qui peuvent être calculées de manière robuste, et fournissent ainsi une mise à l’échelle pour calculer les valeurs propres classiques.

Une seconde partie du travail (collaboration avec L. Grigori) porte sur le calcul de mises-à-l’échelle issues de la résolution d’un problème d’affectation optimale. On a développé un algorithme dont l’idée est de voir le problème d’affectation comme une limite d’un problème de maximisation d’entropie. Ceci conduit à un préprocessing parallèle, qui permet d’éliminer a priori des coefficients qui ne participent pas aux affectations optimales, de sorte que le problème réduit devient résoluble sur une machine séquentielle. L’algorithme ainsi obtenu est étudié dans le preprint [69], qui comprend également des résultats expérimentaux.

\textbf{English version}

The PhD work of M. Sharify [13] deals with the development of scaling methods in matrix analysis to improve the accuracy of numerical computations.

A first part of the work, applying the techniques of [73], [74], deals with eigenvalue problems. We show in particular that the order of magnitude of the eigenvalues of a matrix pencil can be determined (under nondegeracy conditions) by computing tropical eigenvalues. The latter can always be computed accurately and provide a scaling which can be combined with standard numerical methods for matrix pencils.

A second part of the work (collaboration with L. Grigori) deals with the parallel computation of scalings based on the optimal assignment problem. The latter is thought of as a limit of an entropy maximization problem. This leads to a parallel preprocessing, allowing one to eliminate a priori entries which do not belong to optimal assignment, so that the reduced problem becomes solvable on a sequential machine. This algorithm is studied in the preprint [69], which also comprises experimental results.
6.3.3. Mesures et applications maxitives

Participants: Marianne Akian, Paul Poncet.

Les mesures et intégrales maxitives qui ont été introduites et ré-introduites sous divers noms dans la littérature (intégrale de Shilkret, sup-mesures, mesures de possibilité, mesures idempotentes de Maslov, etc.), sont définies de manière analogue aux mesures et intégrales usuelles, en remplaçant les lois additive et multiplicative par celles d’un semi-anneau idempotent, comme par exemple le semi-anneau max-plus. Elles peuvent aussi être obtenues comme limites de mesures positives après déformation logarithmique, par le principe des grandes déviations. Entre autres motivations à l’étude de ces mesures, citons les processus max-stables et leur représentation intégrale, les processus extrêmes, ou les grandes déviations à la loi des grands nombres.

Le travail de thèse de Paul Poncet [12] est parti de ces motivations. Il traite essentiellement de ce que l’on appelle l’analyse idempotente, c’est-à-dire l’étude des espaces fonctionnels ou linéaires de dimension infinie sur l’algèbre tropicale, ou tout autre semi-anneau idempotent. Paul Poncet a développé pour cela un point de vue treillis continu comme dans [1], ou plus généralement domaines, et ses travaux pourraient donc aussi avoir des applications en informatique.

La première partie de la thèse traite des mesures maxitives. Paul Poncet a donné une revue des résultats existants concernant l’existence d’une densité cardinale ou d’une densité d’une mesure par rapport à une autre (théorème de Radon-Nikodym), et la régularité d’une mesure maxitive, tout en les comparant et les complétant. En particulier il prouve une réciproque au théorème de Radon-Nikodym pour les mesures maxitives, c’est-à-dire qu’il donne une caractérisation des mesures maxitives ayant la propriété de Radon-Nikodym, il caractérise les mesures maxitives régulières à valeurs dans un domaine, donne un théorème de décomposition des mesures maxitives aussi publié dans [34], et donne un théorème de représentation de Riesz pour les formes linéaires max-plus continues.

Une deuxième partie concerne les convexes dans les semi-treillis ou l’algèbre max-plus. Paul Poncet s’est intéressé à l’existence d’un théorème de type Krein-Milman, à sa réciproque de Milman, et à celle d’un théorème de type représentation de Choquet dans ces structures. Dans le cas des semi-treillis, certains de ces résultats se déduisent rapidement des travaux sur les semi-treillis compacts, mais d’autres sont entièrement nouveaux. Le théorème de Krein-Milman pour les convexes tropicaux, qui n’avait été établi dans la littérature qu’en dimension finie [137], [96], [125], est prouvé en dimension infinie au moyen de celui sur les semi-treillis. Le théorème de représentation de Choquet utilise les notions de mesures maxitives introduites dans la première partie. De tels résultats permettent de retrouver partiellement les résultats sur la frontière de Martin max-plus décrits dans la section 6.1.1.

Enfin dans une troisième et dernière partie, Paul Poncet étudie les semi-groupes inverses dans une tentative d’unification de l’algèbre usuelle et de l’algèbre tropicale.

English version

Maxitive measures and integrals, which have been introduced and re-introduced under different names in the literature (Shilkret integral, sup-measures, possibility measures, Maslov idempotent measures, etc.), are defined analogously to usual measures and integrals, by replacing the additive and multiplicative laws by the laws of an idempotent semiring, such as the max-plus semiring. They can also be obtained as limits of positive measures after logarithmic deformation, by the large deviation principle. Among motivations for the study of this notion, let us mention max-stable processes and their integral representations, extremal processes, or large deviations to the law of large numbers.

The PhD thesis work of Paul Poncet [12] started from these motivations. It concerns essentially what is called idempotent analysis, that is the study of infinite dimensional functional or linear spaces over tropical algebra, or any other idempotent semiring. For this aim, Paul Poncet developed the point of view of continuous lattices, as in [1], or more generally of domains, and his works may have applications in computer science.
The first part of his thesis concerns maxitive measures. Paul Poncet gave a survey of existing results concerning the existence of a cardinal density of a measure, that of a density of a measure with respect to another (Radon-Nikodym theorem), and the regularity of a maxitive measure, while comparing and extending them. In particular he proves a converse to the Radon-Nikodym theorem for maxitive measures, which lead to a characterisation of maxitive measures that have the Radon-Nikodym property, he characterizes domain valued maxitive measures that are regular, gives a decomposition theorem of maxitive measures also published in [34], and gives a Riesz representation theorem for continuous max-plus linear forms.

A second part concerns convex sets in lattices or max-plus algebra. Paul Poncet is showing a Krein-Milman type theorem, its Milman converse, or a Choquet representation type theorem in these structures. In the case of semilattices, some of these results can be deduced easily from works on compact semilattices, but some others are new. The Krein-Milman on tropical convex sets, which in the litterature was established in finite dimension only [137], [96], [125], is deduced in infinite dimension from the analogous result concerning semilattices. The Choquet representation theorem uses the notions of maxitive measures introduced in the first part. Such results lead in particular to new proofs of some of the results on Martin boundaries described in Section 6.1.1.

In the third and last part, Paul Poncet is studying inverse semigroups in an attempt to unify usual and tropical algebras.

6.4. Algorithmes/Algorithms

6.4.1. Méthodes multigrilles pour le contrôle stochastique et les jeux répétés à somme nulle/Multigrid methods for stochastic control and repeated zero sum games

Participants: Marianne Akian, Sylvie Detournay.

L’algorithme d’itération sur les politiques est bien connu pour résoudre efficacement les équations de la programmation dynamique associées à des problèmes de contrôle stochastique avec critère à horizon infini (Howard) ou ergodique (Denardo et Fox). Récemment, il a été généralisé au cas de problèmes de jeux à deux joueurs et somme nulle dégénérés (avec paiements ergodiques et de type “multi-chaîne”), au moyen de techniques d’algèbre max-plus et de théorie du potentiel non linéaire [103]. Chaque itération de base de cet algorithme utilise la résolution d’un système d’équations linéaires dont l’opérateur est monotone, mais dont la taille peut être grande, soit parce qu’il provient d’une discrétisation fine d’une équation aux dérivées partielles, soit parce qu’il est associé à un problème discret de grande taille comme le graphe du Web.

Or, la méthode multigrille est l’une des rares méthodes permettant de résoudre, au moins dans les bons cas, des systèmes linéaires en un temps de l’ordre de la taille du système. De plus, alors que la méthode multigrille classique ne s’applique qu’à des discrétisations d’équations aux dérivées partielles elliptiques, la méthode multigrille algébrique (voir par exemple [172]) peut s’appliquer à tout système linéaire présentant des propriétés de monotonie (principe du maximum ou système avec M-matrice).

L’association entre méthodes multigrilles et itérations sur les politiques a déjà été utilisée et étudiée dans le cas de problèmes de contrôle stochastique actualisé (voir par exemple [72], [80]), ainsi que dans le cas d’un algorithme d’itération sur les politiques simplifié pour le contrôle ergodique (voir par exemple [5]), mais pour lequel il n’existe pas de preuve de convergence. La méthode multigrille algébrique a été récemment associée à des méthodes d’apprentissage (voir par exemple [187]). Nous l’avons aussi testée dans le cas de l’itération sur les politiques pour des problèmes de jeux à somme nulle actualisés au cours du stage de Shantanu Gangal en 2007.

La thèse de Sylvie Detournay a pour but de développer et d’étudier un algorithme associant une méthode d’itération sur les politiques du type celle introduite par Cochet-Terrasson et Gaubert dans [103] et une méthode multigrille algébrique, afin de résoudre des problèmes de jeux à somme nulle dégénérés, éventuellement posés directement sous forme discrète. Sylvie Detournay a d’abord travaillé sur le cas non dégénéré (actualisé) en codant d’abord seulement l’itération sur les politiques (en C) et appelant des codes libres de méthodes multigrilles algébriques. Ces codes n’étant pas assez souples pour être modifiés, elle a
ensuite codé elle-même certains types de méthodes multigrilles algébriques. Des tests sur des discrétisations d’équations aux dérivées partielles d’Hamilton-Jacobi-Bellman ou d’Isaacs, ou d’inéquations variationnelles ont donné de bons résultats et sont présentés dans [15].

Sylvie Detournay a travaillé cette année sur le cas de problèmes avec critère moyen en temps. Elle a implémenté et raffiné l’algorithme proposé par Cochet-Terrasson et Gaubert [103], en l’associant soit à des méthodes de résolution exacte de systèmes linéaires, soit à des méthodes multigrilles algébriques, en utilisant aussi des méthodes multigrilles multiplicatives pour le calcul de la mesure invariante de chaînes de Markov irréductibles. Ceci a permis en particulier l’obtention de résultats numériques dans le cas de discrétisations d’équations d’Isaacs associées à des jeux de poursuite déterministes ou aléatoires. Plusieurs de ces résultats ont été présentés cette année lors de 2 conférences internationales [44], [45], et devraient faire l’objet d’un article en préparation. Par ailleurs dans un article avec Jean Cochet-Terrasson et Stéphane Gaubert [54], nous présentons l’algorithme, sa convergence et des résultats numériques obtenus avec des méthodes de résolution exacte de systèmes linéaires.

English version

Policy iteration is a powerful and well known algorithm to solve the dynamic programming equation associated to one player problems. It has recently been extended to degenerate two players problems (with ergodic payoff and in “multichain” cases) using ideas from max-plus algebra and nonlinear potential theory [103]. One basic iteration of the algorithm consists in solving a linear system which operator is monotone, but which size may be large since it comes from the discretization of a partial differential equation or since it is associated to a large size discrete problem such as the Web graph.

For the solution of large size linear systems, the state of art consists of multigrid methods which are often able to solve systems in linear time. Whereas multigrid methods can only be applied to systems that come from discretizations of elliptic partial differential equations, algebraic multigrid methods (see for instance [172]) can be applied to any linear system with monotonicity properties (discrete maximum principle or system with a M-matrix).

The association of multigrid methods with policy iteration has been used and studied in the case of discounted stochastic control problems (see for instance [72], [80]), or in the case of a simplified policy iteration algorithm for ergodic control (see for instance [5]), but for which no proof of convergence is known. Some recent work combines the algebraic multigrid method with learning methods [187]. We have also tested it in the case of policy iterations for discounted zero-sum two-player games, during the internship of Shantanu Gangal in 2007.

The aim of the PhD thesis of Sylvie Detournay is to develop and study an algorithm for degenerate two player games (that may come from a discrete time and finite state space model) combining a policy iteration such as that introduced in [103] and an algebraic multigrid method (AMG). Sylvie Detournay has first worked on the nondegenerate (discounted) case, by coding first the policy iterations (in C) and using free AMG softwares. Since these softwares cannot be modified easily, she has then implemented some types of AMG algorithms (in C). Some tests on discretisations of Hamilton-Jacobi-Bellman or Isaacs partial differential equations or variational inequalities gave good results and are presented in [15].

She has worked this year on the case of problems with mean-payoff criteria. She has implemented and refined the algorithm proposed by Cochet-Terrasson and Gaubert [103], associated either to direct linear solvers, or to the AMG methods already used in the nondegenerate case, and also used multiplicative AMG methods developed in the literature for computing invariant measures of Markov chains. This allows her to obtain numerical results in the case of discretisations of Isaacs equations associated to deterministic or stochastic pursuit games. Several of these results were presented this year in 2 international conferences [44], [45] and are part of an article in preparation. Moreover, in an article with Jean Cochet-Terrasson and Stéphane Gaubert [54], we are presenting the algorithm, its convergence and numerical results obtained with direct linear solvers.

6.4.2. Algorithmique des polyèdres tropicaux/Algorithmics of tropical polyhedra
Participants: Xavier Allamigeon, Stéphane Gaubert, Eric Goubault [CEA].

X. Allamigeon, S. Gaubert, et E. Goubault, ont développé dans [82],[57] plusieurs algorithmes permettant de manipuler des polyèdres tropicaux. Ceux-ci correspondent aux travaux décrits dans §6.2.1. Ils permettent notamment de déterminer les sommets et rayons extrêmes d’un polyèdre tropical défini comme intersection de demi-espaces, ou inversement, de calculer une représentation externe à partir d’un ensemble de générateurs. Ces algorithmes sont implémentés la bibliothèqueTPLib (voir §5.3).

English version

X. Allamigeon, S. Gaubert, and E. Goubault, have developed in [82],[57] algorithms allowing one to manipulate tropical polyhedra. They correspond to the contributions described in §6.2.1. In particular, they can be used to determine the vertices and extreme rays of a tropical polyhedron defined as the intersection of half-spaces, or inversely, to compute an external description from a set of generators. These algorithms are implemented in the library TPLib (see §5.3).

6.4.3. Problèmes d’accessibilité dans les hypergraphes orientés et leur complexité/Reachability problems in directed hypergraphs and their complexity

Participant: Xavier Allamigeon.

Les hypergraphes orientés sont une généralisation des graphes orientés, dans lesquelles chaque arc relie un ensemble de sommets à un autre. Ils jouent un rôle important dans les travaux récents sur la convexité tropicale (voir §6.2.1), puisqu’ils offrent une représentation naturelle des cônes définis sur le sous-semi-anneau booléen $B = \{-\infty, 0\}$.

Dans un travail de X. Allamigeon [56], on étudie la complexité de problèmes d’accessibilité sur les hypergraphes orientés. Nous introduisons un algorithme de complexité presque linéaire permettant de déterminer les composantes fortement connexes terminales (qui n’accèdent à aucune autre composante si ce n’est elles-mêmes) d’un hypergraphe.

Nous établissons également une borne inférieure sur-linéaire sur la taille de la réduction transitive de la relation d’accessibilité dans les hypergraphes. Cela indique que la relation d’accessibilité dans les hypergraphes orientés est combinatoirement plus complexe que celle des graphes orientés. Cela suggère aussi que des problèmes comme le calcul des composantes fortement connexes est plus difficile sur les hypergraphes que sur les graphes. Nous mettons d’ailleurs en évidence une réduction en temps linéaire du problème du calcul des ensembles minimaux dans une famille d’ensembles donnée, vers le problème du calcul de toutes les composantes fortement connexes d’un hypergraphe. Le problème du calcul des ensembles minimaux a été largement étudié dans la littérature [163],[183],[182],[164],[165],[166],[118],[88], et aucune algorithm en temps linéaire n’est connu à ce jour.

English version

Directed hypergraphs are a generalization of directed graphs, in which the tail and the head of the arcs are sets of vertices. It appears that they play an important role in the recent works on tropical convexity (see §6.2.1), since they offer a natural representation of cones defined over the boolean sub-semiring $B = \{-\infty, 0\}$.

In a work of X. Allamigeon [56], we study the complexity of reachability problems on directed hypergraphs. We introduce an almost linear-time algorithm allowing to determine the terminal strongly connected components (a component is said to be terminal when no other component is reachable from it).

We also establish a super-linear lower bound over the size of the transitive reduction of the reachability relation in directed hypergraphs. This indicates that the reachability relation is combinatorially more complex in directed hypergraphs than in directed graphs. This also suggests that reachability problems such as computing all strongly connected components are likely to be harder in hypergraphs than in graphs. Besides, we show that the minimal set problem can be reduced in linear time to the problem of computing all strongly connected components in hypergraphs. The former problem consists in finding all minimal sets among a given family of sets. It has been well studied in the literature [163],[183],[182],[164],[165],[166],[118],[88], and no linear time algorithm is known.
6.4.4. Approximation max-plus de fonctions valeurs/Max-plus approximation of value functions

Participants: Stéphane Gaubert, Zheng Qu, Shanjian Tang [Fudan University, Shanghai], William McEneaney [San Diego University].

La thèse de Zheng Qu, démarrée en septembre 2010, supervisée par S. Gaubert et S. Tang, porte sur le développement de méthodes tropicales en programmation dynamique approchée.

Un problème de base consiste à approcher au mieux la fonction valeur d’un problème de contrôle ou de jeux par le supremum d’un petit nombre de fonctions choisies dans un dictionnaire fixé a priori. Ce problème est abordé dans [43]. À l’aide de résultats de Grüber portant sur l’approximation de corps convexes par des polytopes, on donne tout d’abord une borne montrant le caractère inévitable de la malédiction de la dimension, pour certaines méthodes de type base max-plus, lorsque la fonction valeur est $C^2$ et strictement convexe. Ce résultat montre que ces familles de méthodes sont asymptotiquement coûteuses lorsque la précision requise tend vers 0. Elles permettent cependant d’obtenir rapidement des approximations certifiées d’une précision donnée pas trop petite (dans ce cas, la malédiction de la dimension est absente). On s’intéresse ensuite à un problème algorithmique clé sous-jacent à ces méthodes, qui consiste à éliminer dynamiquement des fonctions redondantes intervenant dans la représentation. On démontre dans [43] que ce problème est équivalent à un problème géométrique de localisation, dans lequel la métrique est non symétrique (de type Bregman). Ceci a permis d’appliquer divers algorithmes de localisation, conduisant à une amélioration de la méthode antérieure [155].

Un autre travail de Zheng Qu porte sur les équations de Riccati généralisées associées à des problèmes de contrôles stochastique avec critère quadratique, dans lesquels la dynamique comporte un terme bilinéaire en le contrôle et le bruit. Alors que le flot de l’équation de Riccati classique est contractant pour la métrique Riemannienne invariante, pour la métrique de Thompson, ainsi que pour toutes les métriques de Finsler invariantes sur le cône des matrices symétriques positives, on montre ici que le flot de l’équation de Riccati généralisée en question est seulement contractant pour la métrique de Thompson (sous des hypothèses naturelles).

English version

The PhD work of Zheng Qu, which started in September 2010, and is supervised by S. Gaubert and S. Tang, aims in particular at developing tropical methods in approximate dynamic programming.

A basic problem consists in approximating the value function of an optimal control or game problem by a supremum of a small number of functions taken from a prescribed dictionary. This problem is addressed in [43]. By applying results of Grüber concerning the approximation of convex bodies by polytopes, we give first a negative result, showing that the curse of dimensionality cannot be avoided by a family of max-plus basis methods, when the value function is $C^2$ and strictly convex. This result shows that this family of methods is asymptotically computationally expensive when the requested precision tends to 0. However, they can be used to obtain quickly (in a curse of dimensionality free way) certified approximations with a fixed (not too small) precision. Then, we address a key algorithmic subproblem, consisting in trimming dynamically the redundant functions in a max-plus representation. We showed in [43] that this problem is equivalent to a geometric facility location problem, with a non symmetric Bregman type metric. This allowed us to apply several facility location algorithms, leading to an improvement of the earlier method [155].

Another work of Zheng Qu deals with the generalized Riccati equations associated to stochastic optimal control problems with quadratic cost, in which the dynamics comprises a term which is bilinear in the control and in the noise. Whereas the flow of the standard Riccati equation is known to be a contraction for the invariant Riemannian metric, the Thompson metric, and more generally, for all invariant Finsler metrics on the cone of positive definite matrices, it is shown here that the flow of this generalized Riccati equation is only contracting with respect to Thompson metric (under natural assumptions).

6.5. Applications
6.5.1. Introduction
Nous présentons maintenant plusieurs travaux de nature appliquée, touchant à des domaines variés, dans lesquels nous exploitons certaines des techniques mathématiques présentées précédemment, et particulièrement celles qui relèvent de la théorie de Perron-Frobenius non-linéaire et de la convexité tropicale. Ces applications utilisent aussi des techniques d’algèbre linéaire ou d’optimisation convexe.

English version
In this section, we describe several applied works in which we use some of the theoretical tools developed by the team, including non-linear Perron-Frobenius theory and tropical convexity. Some of these applications also make an intensive use of linear algebraic and convex programming methods.

6.5.2. Propriétés des valeurs propres de Perron et de Floquet, et application en chronothérapeutique/Properties of Perron and Floquet eigenvalue, with an application to chronotherapeutics
Participants: Frédérique Billy [Projet BANG, INRIA], Jean Clairambault [Projet BANG, INRIA], Olivier Fercoq, Stéphane Gaubert, Thomas Lepoutre [Projet BANG puis DRACULA, INRIA].

On s’intéresse à des modèles de systèmes dynamiques monotones structurés en âge représentant la croissance de populations de cellules (saines ou tumorales), à la suite de travaux de Clairambault et Perthame. Il s’agit de comprendre l’influence du contrôle circadien sur la croissance des cellules. Dans le cas stationnaire, le taux de croissance est représenté par une valeur propre de Perron. Dans le cas périodique, il s’agit d’une valeur propre de Floquet. Les travaux [42], [52], [58] portent sur l’identification de ces modèles ainsi que sur un problème de contrôle thérapeutique, consistant à minimiser le taux de croissance des cellules tumorales sous une contrainte de non-toxicité du traitement (maintien d’une population de cellules saines). Ce travail s’appuie en particulier sur un algorithme d’optimisation de la valeur propre de Perron d’une matrice développé par Fercoq dans un autre contexte [62].

English version
We study monotone dynamical systems representing the growth of cells (healthy or tumoral), following a work of Clairambault and Perthame. The goal is to understand how the circadian control influences the growth of cells. In the case of stationary monotone systems, this growth is measured by the Perron root. In the time periodic case, this Perron root is replaced by a Floquet multiplier.

The works [42], [52], [58] deal with the identification of these models, together with a therapeutic control problem, consisting in minimizing the growth rate of tumoral cells, under a non-toxicity constraint (preserving the population of healthy cells). This works relies in particular on a fast algorithm to optimize the Perron eigenvalue of a matrix, developed by Fercoq in a different context [62].

6.5.3. Équations aux dérivées partielles en dynamique des populations/Partial differential equations from population dynamics
Participant: Sepideh Mirrahimi.

Un des problèmes sur lequel on a travaillé est un modèle de propagation de populations sexuées dans l’espace [64] où on dérive un modèle étudié par des biologistes, à partir d’un modèle de populations structurées et on étudie la dynamique de ce dernier. Nous avons aussi travaillé sur un problème de protéines motrices, où nous étudions le comportement asymptotique d’un système de deux équations couplées de Fokker-Planck dans un environnement périodique. Avec une approche d’homogénéisation et en utilisant des techniques de solutions de viscosité, on montre que les protéines se déplacent dans une direction constante [65]. De plus, en utilisant des idées venant des problèmes similaires mais discrets, avec Stéphane Gaubert nous étudions les EDPs qui décrivent la dynamique d’une population asexuée sous l’effet des mutations et de la sélection naturelle, et on cherche à déterminer la limite en temps long de la densité de population.

English version
One of the problems studied is a model of propagation of a sexual population in space [64], where we derive a model studied by biologists, from a structured population model. We then study the behavior of the solution to the latter model. We also have worked on a problem of motor proteins, where we study the asymptotic behavior of a time dependent, weakly coupled, Fokker-Planck system of two equations set in a periodic environment. By a homogenization approach and using viscosity solutions technics we prove that the molecules either move along a fixed filament with a constant speed and direction or remain immobile [65]. Moreover, using ideas coming from discrete models, with Stéphane Gaubert we study some PDEs that describe the dynamics of an asexual population under mutations and natural selection. We try to determine the long-time limit of the population density.

6.5.4. Identification du trafic dans les réseaux IP/Traffic identification in IP networks
Participants: Mustapha Bouhtou [Orange Labs], Stéphane Gaubert, Guillaume Sagnol.

Le travail de thèse de Guillaume Sagnol, réalisé en collaboration avec Orange Labs dans le cadre d’un contrat “CRE”, a porté sur l’identification du trafic dans des réseaux IP, problème auquel il a appliqué des d’optimisation SDP et d’optimisation sous-modulaire afin de développer des algorithmes passant à l’échelle. Cette thèse s’est achevée fin 2010 [173]. Les articles suivants relatifs au travail de thèse ont été publiés cette année: [36], [35].

English version
The PhD work of Guillaume Sagnol, done in collaboration with Orange Labs in the framework of a “CRE” research contract, dealt with the identification of the traffic in IP networks. Sagnol applied SDP and submodular optimization techniques to develop scalable algorithms (adapted to large networks). The PhD defense took place in 2010 [173]. Some contributions of the Phd have been published this year in [36], [35].

6.5.5. Analyse statique de programmes et itération sur les politiques/Static analysis of computer programs and policy iteration
Participants: Assale Adjé, Stéphane Gaubert, Eric Goubault [CEA].


Un problème important consiste à déterminer le plus petit point fixe (l’algorithme de [14] fournit un point fixe, qui peut ne pas être minimal). Ce problème est abordé dans [30], où l’approche de [14] est comparée avec une approche duale développée par Gawlitza et Seidl.

English version
The PhD work of A. Adjé [11], co-supervised by S. Gaubert and E. Goubault, applies methods from game theory and optimization (generalized duality, convex and non convex programming) to the fixed point problems arising in static analysis of programs by abstract interpretation. We introduced in [14] a new domain in static analysis, which extends to nonlinear cases the “templates” introduced by Manna, Sankaranarayanan, and Sipma [175]. This domain allows one to represent accessible sets that are non convex. These are defined by finitely many inequalities taken from a dictionary. This allows one to use in particular the information provided by Lyapunov functions, which are often known in applications arising from engineering. We showed in [14] that experimentally accurate invariants can be obtained by coupling policy iteration with Shor.
relaxation (SDP relaxation of convex programming problems). This yields accurate abstractions of some numerical programs, like linear filters with thresholds.

An important problem consists in determining the smallest fixed point (the algorithm of [14] yields a possibly non minimal fixed point). This problem is addressed in [30], in which the approach of [14] is compared with a dual approach developed by Gawlitza and Seidl.

6.5.6. Optimisation du référencement sur la toile/Optimization of web referencing

Participants: Marianne Akian, Mustapha Bouhtou [Orange Labs], Olivier Fercoq, Stéphane Gaubert.

La thèse d’O. Fercoq, co-encadrée par M. Akian, M. Bouhtou, et S. Gaubert, financée par un CRE d’Orange Labs, a pour but d’appliquer des méthodes d’optimisation et de théorie des jeux à l’optimisation de services en lignes. On a tout d’abord étudié le problème de l’optimisation du référencement, que l’on formalise en se donnant par exemple un ensemble d’hyperliens et de ressources obligatoires, dont la nature et la position sur le site web sont déterminées à l’avance par le concepteur. Cet ensemble forme en quelque sorte le squelette du site web. On se donne aussi un ensemble d’hyperliens ou de ressources facultatives, pour lesquels le concepteur du site a certains degrés de liberté (le lien ou le contenu peut être mis sur une page plutôt qu’une autre, voire être omis).

Dans [61], on aborde le problème de l’optimisation du “Pagerank” dans ce cadre, en appliquant des techniques de décision Markovienne classiques et sous-contraintes. Le problème peut en effet se ramener à un problème de contrôle ergodique ou de contrôle ergodique sous contraintes (ergodiques), selon que les contraintes sur les hyperliens sont locales à chaque page ou font intervenir plusieurs pages. On traite à la fois le cas relaxé où les probabilités de passage d’une page à une autre peuvent être des réels positifs quelconques (on peut par exemple supposer que cette probabilité dépend de la position et des caractères utilisés pour l’hyperlien correspondant) et le cas discret où ces probabilités sont uniformes parmis celles qui sont strictement positives (comme dans la modélisation classique conduisant au calcul du Pagerank). On montre que cette famille de problèmes correspondent à des problèmes de programmation dynamique avec un nombre exponentiel de contrôles, mais où les polytopes des mesures de probabilités de transition admettent des oracles de séparation polynômiaux. On obtient de la sorte des résultats de complexité, ainsi que, sous certaines hypothèses, des algorithmes adaptés à des instances de grande taille, couplant programmation dynamique et relaxation Lagrangienne. Ces algorithmes ont été testés sur un fragment du graphe du web.

Un critère de référencement classique, alternatif au pagerank, est donné par le vecteur propre de Perron. O. Fercoq a abordé le problème associé d’optimisation du référencement, qui se révèle plus difficile que celui du pagerank, en raison de l’absence de propriété de convexité. Cependant, il a développé un algorithme rapide et creux (basé sur des propriétés de rang 1 d’opérateurs intervenant dans le calcul de dérivées du critère) permettant de calculer un optimum local du référencement [62]. Il a enfin donné un algorithme analogue pour optimiser le score “HOTS” de Tomlin.

English version

The goal of the PhD work of O. Fercoq, cosupervised by M. Akian, M. Bouhtou, and S. Gaubert, and supported by a research contract (CRE) of Orange Labs, is to apply optimization and game theory methods to the optimization of online services. We started by investigating the problem of the optimization of referencing, which we modelled by considering a family of compulsory hyperlinks and resources (fixed in advance by the website designer, these constitute the “skeletton” of the website) and also a family of facultative hyperlink or resources (some links may be ommited or some other links may be added).

In [61], we are approaching the problem of the pagerank optimization in this framework, by applying usual and constrained Markov decision processes techniques. This problem can indeed be reduced to an ergodic control problem without or with (ergodic) constraints, depending on the fact that hyperlinks constraints are local to each web page or depend on several web pages. We study the relaxed problem where the transition probabilities from one page to another may be any positive real (one may assume for instance that this probability depends on the position and type used for the corresponding hyperlink), as well as the discrete problem where these probabilities are uniform among the positive ones (as in the usual modelisation leading
to the Pagerank). We show that these problems can be reduced to dynamic programming problems with exponentially many discrete actions, in which however the polytopes of transition probability measures admit polynomial time separation oracles. We derive from this approach polynomial time complexity results, as well as under some additional assumption, scalable algorithms (adapted to large web graphs), coupling dynamic programming and Lagrange relaxation. The latter have been tested on a real subgraph of the web.

A classical alternative ranking relies on the Perron eigenvector. O. Fercoq treated the associated optimisation problem, which turns out to be harder than in the pagerank case, due to the lack of convexity properties. However, he developed a fast (sparse) algorithm, exploiting the rank 1 properties of operators appearing when computing the derivative of the objective function, allowing one to compute a local optimum [62]. He also developed a similar method to optimize Tomlin’s “HOTS” score.

6.5.7. Gestion du revenu appliquée à la tarification de services données/Yield management applied to pricing of data services

Participants: Mustapha Bouhtou [Orange Labs], Jean-Baptiste Dumont, Stéphane Gaubert.

Le travail de thèse CIFRE de J-B. Dumont, qui a démarré en Septembre 2010, sous la supervision de M. Bouhtou et S. Gaubert, porte sur la tarification de services data et la gestion des ressources dans les réseaux mobiles, qui est abordée à l’aide de techniques de contrôle et d’optimisation stochastique. Dumont a développé un premier modèle de tarification, permettant d’inciter les clients à reporter leur demande en dehors des périodes les plus chargées.

English version

The CIFRE PhD work of J-B. Dumont started in September 2012, under the joint supervision of M. Bouhtou and S. Gaubert. It deals with the pricing of data services and resource allocation in mobile networks. This is addressed through stochastic control and stochastic optimization techniques. Dumont developed a first model of pricing, giving some incentive to the customers to move their demand from loaded to less loaded time periods.
6. New Results

6.1. Verification of Concurrent Recursive Programs

We have introduced a mathematical model to capture the behavior of concurrent and recursive systems. We have also identified typical properties of these systems that programmers may want to verify. We have come up with a specification language which is powerful enough to express such properties. In fact, we give a framework by which programmers can define their own specification language depending on the specific application as long as the semantics of the operators can be defined in monadic second-order logic. We have shown that checking whether a specified property is satisfiable or whether a given system satisfies a property specified in such a language is decidable with a manageable complexity (double exponential time). The proof technique is so general that it captures the results for various other well studied models. Our results were presented at MFCS’11 [50].

6.2. Product form for Petri nets

While product-form Petri nets have been intensively studied some important questions were left open. In [52], we have solved most of the open problems. We have provided a sound and complete set of rules to synthesise product form Petri nets. We have characterized the complexity class of standard problems (liveness, reachability and cover ability). At last we have proposed a large subclass of product form Petri nets for which the normalising constant (a key quantity) can be efficiently computed. This paper has obtained the outstanding paper award of the ATPN’2011 conference.

6.3. Statistical model checking

We have designed a logic HASL for stochastic systems that can express elaborated performance indices related to path behaviours [46]. We have shown how it can be integrated in the design process of flexible manufacturing systems [47]. We have developed a tool COSMOS for statistical model checking of HASL formula over a stochastic Petri net with general distributions [45]. In parallel, we have developed the first importance sampling method for rare event that still produces a confidence interval (rather than an estimated value) and we have integrated this method in COSMOS [48].

6.4. Symmetry and stochastic systems

We have developed a framework to efficiently solve large Markov decision problems specified by high-level Petri nets [42]. Such a specification allows to decrease the size of the MDP by the analysis of the symmetries of the model. In a purely probabilistic context, we have designed two complementary methods for handling partial symmetries in stochastic high level Petri nets and studied their efficiency on several relevant case studies [41].

6.5. Automata for Data Words

We studied data words, i.e, strings where each position carries both a label from a finite alphabet and some values from an infinite domain. Data words are suitable to model the behavior of concurrent systems with dynamic process creation, as the infinite alphabet can be used to represent an unbounded number of process identifiers. A variety of formalisms, including logic and automata, have been studied to specify sets of data words in the context of verification. However, logic and automata that capture dynamic communicating systems were missing. We closed this gap and developed a quite general logical and automata-theoretic framework for the specification and implementation of sets of data words. On the specification side, we considered a fragment of monadic second-order (MSO) logic, which comes with a predicate to test two word
positions for data equality. As a model of an implementation, we introduced class register automata. Our model combines the well known models of register automata and class memory automata, and it indeed captures dynamic communicating automata, whose semantics can be described as a set of message sequence charts. We studied the realizability problem and show that every formula from the existential fragment of MSO logic can be effectively translated into a class register automaton. These results were presented at CONCUR'11 [49].

6.6. Weighted Logics with Navigation for Trees

We continued our study of the verification of quantitative properties and applications to queries over XML documents. Verification of quantitative systems follow a classical scheme in three steps: specification, modeling, and algorithmics. Hence, we started by exhibiting a specification language. To describe natural qualitative properties, we chose to use, as a fragment, boolean logic like first-order logic or monadic second-order logic. We then encapsulate this properties into the quantitative formalism, allowing sums and products computations in a specified general semiring. In the word case, we obtained very strong results relating this kind of specification/computation languages with the well-known weighted finite automata, and the new weighted pebble automata, which permit to model several interesting quantitative computations over words. We extended these results to trees, and in particular, finite unranked trees or nested words, which are a natural model for XML documents. We published preliminary results in a research report [57], and we have worked on a submission of these results to several conferences. Our next goal is to tackle some of the algorithmic questions that naturally arise in this context, like satisfiability or model checking.

6.7. Contextual Petri nets

Contextual nets are an extension of Petri nets that – unlike ordinary Petri nets – faithfully models concurrent read accesses to shared resources. This is not only interesting from a semantic but also from an algorithmic point of view, as the analysis of such nets can better exploit the fact that concurrent reads are independent and concurrent.

In particular, the unfolding of a contextual net may be up to exponentially smaller in certain situations. While the theoretical foundations of contextual unfoldings were established in [66] and [6], it remained unclear whether the approach could be useful in practice.

Recent work on the theoretical foundations, as well as appropriate data structures and algorithms, has closed this gap. This underlying work has been presented at Concur’11 [53] and has resulted in an efficient tool [59]. More details can be found in a technical report [58].

We are currently exploring applications of these techniques in the areas of verification, diagnosis, and planning. Some preliminary steps have been presented in [54].

6.8. Occurrence net Synthesis

Occurrence nets are a well known partial order model for the concurrent behavior of Petri nets. The causality and conflict relations between events, which are explicitly represented in occurrence nets, induce logical dependencies between event occurrences: the occurrence of an event e in a run implies that all its causal predecessors also occur, and that no event in conflict with e occurs. But these structural relations do not express all the logical dependencies between event occurrences in maximal runs: in particular, the occurrence of e in any maximal run may imply the occurrence of another event that is not a causal predecessor of e, in that run. The reveals relation had been introduced in [33] to express this dependency between two events. In this work, presented at ACSD 2011 [44], we extend the theory in two ways: first, we generalize the reveals relation to express more general dependencies, involving more than two events, and we introduce ERL logic to express these dependencies as boolean formulas. Secondly, we solve the synthesis problem that arises: given an ERL formula φ, is there a finite occurrence net N such that φ describes exactly the dependencies between the events of N? The resulting method requires only two synthesis rules.
6.9. Computing Reveals

The reveals relation has been introduced in [33] between events in occurrence nets. Essentially, event $a$ is said to reveal event $b$ if in any maximal run that contains $a$, $b$ must also occur, be it before, after, or concurrently with $a$, and even if $a$ and $b$ are not causally related. Information of this kind is useful for diagnosis; a sensor for event $a$ may render a sensor for $b$ unnecessary.

In [33], the reveals relation was shown to be decidable for occurrence nets that represent unfoldings of safe Petri nets. However, the upper bound was prohibitive for computing the relation in practice. In [51] we address this problem and drastically reduce the upper bound. We also propose efficient algorithms to actually compute the relation on a given occurrence net.
6. New Results

6.1. A supervised clustering approach for fMRI-based inference of brain states

We propose a method that combines signals from many brain regions observed in functional Magnetic Resonance Imaging (fMRI) to predict the subject’s behavior during a scanning session. Such predictions suffer from the huge number of brain regions sampled on the voxel grid of standard fMRI data sets: the curse of dimensionality. Dimensionality reduction is thus needed, but it is often performed using a univariate feature selection procedure, that handles neither the spatial structure of the images, nor the multivariate nature of the signal. By introducing a hierarchical clustering of the brain volume that incorporates connectivity constraints, we reduce the span of the possible spatial configurations to a single tree of nested regions tailored to the signal. We then prune the tree in a supervised setting, hence the name supervised clustering, in order to extract a parcellation (division of the volume) such that parcel-based signal averages best predict the target information. Dimensionality reduction is thus achieved by feature agglomeration, and the constructed features now provide a multi-scale representation of the signal. Comparisons with reference methods on both simulated and real data show that our approach yields higher prediction accuracy than standard voxel-based approaches. Moreover, the method infers an explicit weighting of the regions involved in the regression or classification task. See also [14] and Fig. 1.

Figure 1. Results for prediction of object size. Maps of weights found by supervised cut in the prediction of the size of an object. The proposed algorithm creates very interpretable clusters, compared to the reference methods that do not consider the spatial structure of the image.

6.2. Multiclass Sparse Bayesian Regression for fMRI-Based Prediction

Inverse inference has recently become a popular approach for analyzing neuroimaging data, by quantifying the amount of information contained in brain images on perceptual, cognitive, and behavioral parameters. As it outlines brain regions that convey information for an accurate prediction of the parameter of interest, it allows to understand how the corresponding information is encoded in the brain. However, it relies on a prediction function that is plagued by the curse of dimensionality, as there are far more features (voxels) than samples (images), and dimension reduction is thus a mandatory step. We introduce in this work a new model, called Multiclass Sparse Bayesian Regression (MCBR), that, unlike classical alternatives, automatically adapts the amount of regularization to the available data. MCBR consists in grouping features into several classes and then regularizing each class differently in order to apply an adaptive and efficient regularization. We detail these framework and validate our algorithm on simulated and real neuroimaging data sets, showing that it performs better than reference methods while yielding interpretable clusters of features. See also [13] and Fig. 2.
Figure 2. Mental representation of size - Inter-subject analysis. Histogram of the weights found by Gibbs-MCBR, and corresponding class membership values (each color of dots represents a different class), for the inter-subject analyzes on the mental representation of size. We can see that Gibbs-MCBR creates clusters of informative and non-informative voxels, and that the different classes are regularized differently, according to the relevance of the features within them.

6.3. Total variation regularization for fMRI-based prediction of behaviour

While medical imaging typically provides massive amounts of data, the extraction of relevant information for predictive diagnosis remains a difficult challenge. Functional MRI (fMRI) data, that provide an indirect measure of task related or spontaneous neuronal activity, are classically analyzed in a mass-univariate procedure yielding statistical parametric maps. This analysis framework disregards some important principles of brain organization: population coding, distributed and overlapping representations. Multivariate pattern analysis, i.e., the prediction of behavioural variables from brain activation patterns better captures this structure. To cope with the high dimensionality of the data, the learning method has to be regularized. However, the spatial structure of the image is not taken into account in standard regularization methods, so that the extracted features are often hard to interpret. More informative and interpretable results can be obtained with the $\ell_1$ norm of the image gradient, a.k.a. its Total Variation (TV), as regularization. We apply for the first time this method to fMRI data, and show that TV regularization is well suited to the purpose of brain mapping while being a powerful tool for brain decoding. Moreover, this article presents the first use of TV regularization for classification. See also [15] and Fig. 3.

6.4. Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom.

As it provides the only method for mapping white matter fibers in vivo, diffusion MRI tractography is gaining importance in clinical and neuroscience research. However, despite the increasing availability of different diffusion models and tractography algorithms, it remains unclear how to select the optimal fiber reconstruction method, given certain imaging parameters. Consequently, it is of utmost importance to have a quantitative comparison of these models and algorithms and a deeper understanding of the corresponding strengths and weaknesses. In this work, we use a common dataset with known ground truth and a reproducible methodology to quantitatively evaluate the performance of various diffusion models and tractography algorithms. To examine a wide range of methods, the dataset, but not the ground truth, was released to the public for evaluation in a contest, the "Fiber Cup". 10 fiber reconstruction methods were evaluated. The results provide evidence
Figure 3. Regression - Sizes prediction experiment - Inter-subject analysis. Maps of weights found by TV regression for various values of the regularization parameter \( \lambda \). When \( \lambda \) decreases, the TV regression algorithm creates different clusters of weights with constant values. These clusters are easily interpretable, compared to voxel-based map (see below). The TV regression algorithm is very stable for different values of \( \lambda \).
that: 1. For high SNR datasets, diffusion models such as (fiber) orientation distribution functions correctly model the underlying fiber distribution and can be used in conjunction with streamline tractography, and 2. For medium or low SNR datasets, a prior on the spatial smoothness of either the diffusion model or the fibers is recommended for correct modelling of the fiber distribution and proper tractography results. The phantom dataset, the ground truth fibers, the evaluation methodology and the results obtained so far will remain publicly available on [http://www.lnao.fr](http://www.lnao.fr) See also [10].

6.5. Multi-subject dictionary learning (MSDL) to segment an atlas of brain spontaneous activity

Fluctuations in brain on-going activity can be used to reveal its intrinsic functional organization. To mine this information, we give a new hierarchical probabilistic model for brain activity patterns that does not require an experimental design to be specified. We estimate this model in the dictionary learning framework, learning simultaneously latent spatial maps and the corresponding brain activity time-series. Unlike previous dictionary learning frameworks, we introduce an explicit difference between subject-level spatial maps and their corresponding population-level maps, forming an atlas. We give a novel algorithm using convex optimization techniques to solve efficiently this problem with non-smooth penalties well-suited to image denoising. We show on simulated data that it can recover population-level maps as well as subject specificities. On resting-state fMRI data, we extract the first atlas of spontaneous brain activity and show how it defines a subject-specific functional parcellation of the brain in localized regions. See also [25] and Fif 4.

![Figure 4. Outlines at 33% of all dictionary elements estimated by MSDL for 2 different set of cutting planes. The motor system is divided in (1) dorsal, (2) lateral, and (3) ventral regions. Similarly, the visual system is divided in (4) a primary region centered on the Calcarine sulcus, overlapping with (5) a region centered on the striate cortex, and (6) extrastriate regions. (7), (8): fine details of the vascular system segmented in several maps.](image)

6.6. Functional brain imaging with M/EEG using structured sparsity in time-frequency dictionaries
Magnetoencephalography (MEG) and electroencephalography (EEG) allow functional brain imaging with high temporal resolution. While time-frequency analysis is often used in the field, it is not commonly employed in the context of the ill-posed inverse problem that maps the MEG and EEG measurements to the source space in the brain. In this work, we detail how convex structured sparsity can be exploited to achieve a principled and more accurate functional imaging approach. Importantly, time-frequency dictionaries can capture the non-stationary nature of brain signals and state-of-the-art convex optimization procedures based on proximal operators allow the derivation of a fast estimation algorithm. We compare the accuracy of our new method to recently proposed inverse solvers with help of simulations and analysis of real MEG data. See also [22].

6.7. A probabilistic framework to infer brain functional connectivity from anatomical connections

We present a novel probabilistic framework to learn across several subjects a mapping from brain anatomical connectivity to functional connectivity, i.e. the covariance structure of brain activity. This prediction problem must be formulated as a structured-output learning task, as the predicted parameters are strongly correlated. We introduce a model selection framework based on cross-validation with a parametrization-independent loss function suitable to the manifold of covariance matrices. Our model is based on constraining the conditional independence structure of functional activity by the anatomical connectivity. Subsequently, we learn a linear predictor of a stationary multivariate autoregressive model. This natural parameterization of functional connectivity also enforces the positive-definiteness of the predicted covariance and thus matches the structure of the output space. Our results show that functional connectivity can be explained by anatomical connectivity on a rigorous statistical basis, and that a proper model of functional connectivity is essential to assess this link. See also [20] and Fig. 5.

Figure 5. Identifying structural connections associated with the default mode network. With yellow is represented the lateral parietal cortex, green areas represent the posterior cingulate gyrus (PCC), blue and light blue represent the medial prefrontal and orbito-frontal areas, respectively. The right model performs much better in terms of cross-validated data likelihood.

6.8. M/EEG source reconstruction based on Gabor thresholding in the source space

Thanks to their high temporal resolution, source reconstruction based on Magnetoencephalography (MEG) and/or Electroencephalography (EEG) is an important tool for noninvasive functional brain imaging. Since the MEG/EEG inverse problem is ill-posed, inverse solvers employ priors on the sources. While priors are generally applied in the time domain, the time-frequency (TF) characteristics of brain signals are rarely employed as a spatio-temporal prior. In this work, we present an inverse solver which employs a structured sparse prior formed by the sum of $\ell_{21}$ and $\ell_1$ norms on the coefficients of the Gabor TF decomposition of the
source activations. The resulting convex optimization problem is solved using a first-order scheme based on proximal operators. We provide empirical evidence based on EEG simulations that the proposed method is able to recover neural activations that are spatially sparse, temporally smooth and non-stationary. We compare our approach to alternative solvers based also on convex sparse priors, and demonstrate the benefit of promoting sparse Gabor decompositions via a mathematically principled iterative thresholding procedure. See also [24].

6.9. Multifractal Analysis of Resting State Networks in Functional MRI

It has been known for at least one decade that functional MRI time series display long-memory properties, such as power-law scaling in the frequency spectrum. Concomitantly, multivariate model-free analysis of spatial patterns, such as spatial Independent Component Analysis (sICA), has been successfully used to segment from spontaneous activity Resting-State Networks (RSN) that correspond to known brain function. As recent neuroscientific studies suggest a link between spectral properties of brain activity and cognitive processes, a burning question emerges: can temporal scaling properties offer new markers of brain states encoded in these large scale networks? In this work, we combine two recent methodologies: group-level canonical ICA for multi-subject segmentation of brain network, and wavelet leader-based multifractal formalism for the analysis of RSN scaling properties. We identify the brain networks that elicit self-similarity or multifractality and explore which spectral properties correspond specifically to known functionally relevant processes in spontaneous activity. See also [19].

6.10. Multi-scale Mining of fMRI Data with Hierarchical Structured Sparsity

Inverse inference, or “brain reading”, is a recent paradigm for analyzing functional magnetic resonance imaging (fMRI) data, based on pattern recognition tools. By predicting some cognitive variables related to brain activation maps, this approach aims at decoding brain activity. Inverse inference takes into account the multivariate information between voxels and is currently the only way to assess how precisely some cognitive information is encoded by the activity of neural populations within the whole brain. However, it relies on a prediction function that is plagued by the curse of dimensionality, as we have far more features than samples, i.e., more voxels than fMRI volumes. To address this problem, different methods have been proposed. Among them are univariate feature selection, feature agglomeration and regularization techniques. In this work, we consider a hierarchical structured regularization. Specifically, the penalization we use is constructed from a tree that is obtained by spatially constrained agglomerative clustering. This approach encodes the spatial prior information in the regularization process, which makes the overall prediction procedure more robust to inter-subject variability. We test our algorithm on a real data acquired for studying the mental representation of objects, and we show that the proposed algorithm yields better prediction accuracy than reference methods. See also [29] and Fig. 6.

6.11. Detecting Outlying Subjects in High-Dimensional Neuroimaging Datasets with Regularized Minimum Covariance Determinant

Medical imaging datasets used in clinical studies or basic research often comprise highly variable multi-subject data. Statistically-controlled inclusion of a subject in a group study, i.e. deciding whether its images should be considered as samples from a given population or whether they should be rejected as outlier data, is a challenging issue. While the informal approaches often used do not provide any statistical assessment that a given dataset is indeed an outlier, traditional statistical procedures are not well-suited to the noisy, high-dimensional, settings encountered in medical imaging, e.g. with functional brain images. In this work, we modify the classical Minimum Covariance Determinant approach by adding a regularization term, that ensures that the estimation is well-posed in high-dimensional settings and in the presence of many outliers. We show on simulated and real data that outliers can be detected satisfactorily, even in situations where the number of dimensions of the data exceeds the number of observations. See also [21] and Fig. 7.
Figure 6. Principle of structured sparsity: Example of a tree $T$ when $p = 5$, with three voxels and two parcels. The parcel 2 is defined as the averaged intensity of the voxels $\{1, 2\}$, while the parcel 1 is obtained by averaging the parcel 2 and voxel 3. In red dashed lines are represented the five groups of variables that compose $\mathcal{S}$. If the group containing the parcel 2 is set to zero, the voxels $\{1, 2\}$ are also (and necessarily) zeroed out.

Figure 7. Regularized-MCD-based Mahalanobis distances of a small sample. The higher the Mahalanobis distance, the higher the probability for an observation to be tagged as outlying. Points in red are outliers subjects according to the whole population.
6.12. Connectivity-informed fMRI Activation Detection

A growing interest has emerged in studying the correlation structure of spontaneous and task-induced brain activity to elucidate the functional architecture of the brain. In particular, functional networks estimated from resting state (RS) data were shown to exhibit high resemblance to those evoked by stimuli. Motivated by these findings, we propose a novel generative model that integrates RS-connectivity and stimulus-evoked responses under a unified analytical framework. Our model permits exact closed-form solutions for both the posterior activation effect estimates and the model evidence. To learn RS networks, graphical LASSO and the oracle approximating shrinkage technique are deployed. On a cohort of 65 subjects, we demonstrate increased sensitivity in fMRI activation detection using our connectivity-informed model over the standard univariate approach. Our results thus provide further evidence for the presence of an intrinsic relationship between brain activity during rest and task, the exploitation of which enables higher detection power in task-driven studies. See also [23] and Fig 8.

![Figure 8. Real data results. (a) rate of parcels with significant activation differences averaged across contrasts vs. p-value thresholds. (b) Parcels detected by contrasting computation against sentence processing task, and (c) auditory against visual task. Red = detected by only OAS-CM. Purple = detected by both OAS-CM and GL-CM. Blue = detected by all methods.](image)

6.13. Beyond brain reading: identify and predict with clustering and randomized sparsity

The prediction of behavioral covariates from functional MRI (fMRI) is known as brain reading. From a statistical standpoint, this challenge is a supervised learning task. The ability to predict cognitive states from new data gives a model selection criterion: prediction accuracy. While a good prediction score implies that some of the voxels used by the classifier are relevant, one cannot state that these voxels form the brain regions involved in the cognitive task. The best predictive model may have selected by chance non-informative regions, and neglected relevant regions providing duplicate information. In this contribution, we address the support identification problem. The proposed approach relies on randomization techniques which have been proved to be consistent for support recovery. To account for the spatial correlations between voxels, our approach makes use of a spatially constrained hierarchical clustering algorithm. Results are provided on simulations and a visual experiment. See Fig. 9.

6.14. Joint T1 and Brain Fiber Diffeomorphic Registration Using the Demons

Non-linear image registration is one of the most challenging tasks in medical image analysis. In this work, we propose an extension of the well-established diffeomorphic Demons registration algorithm to take into
account geometric constraints. Combining the deformation field induced by the image and the geometry, we define a mathematically sound framework to jointly register images and geometric descriptors such as fibers or sulcal lines. We demonstrate this framework by registering simultaneously T1 images and 50 fiber bundles consistently extracted in 12 subjects. Results show the improvement of fibers alignment while maintaining, and sometimes improving image registration. Further comparisons with non-linear T1 and tensor registration demonstrate the superiority of the Geometric Demons over their purely iconic counterparts. See also [28] and Fig. 10.

Figure 9. Results on fMRI object recognition task (face vs. house). The selected voxels are in the Fusiform Face Area. left. Prediction Receiver-Operating Characteristic. right. Scores with Ward Randomized Logistic regression.

Figure 10. Influence of the fiber weighting term on the registration accuracy. Fibers of 11 subjects were overlapped after registration with the Geometric Demons for three values of the fiber weighting parameter. Corresponding fibers in different subjects share colors.
6. New Results

6.1. The Focused Calculus of Structures

Participants: Kaustuv Chaudhuri, Nicolas Guenot, Lutz Straßburger.

The *sequent calculus* is a proof system for logic that has many nice properties from a proof search perspective, the most famous being the subformula property that is essential to tame the search space. In recent years, the *focusing* property of sequent systems has become another useful property, both for shrinking the search space and to improve the representation of proofs. However, the sequent calculus does have some limitations. Primarily, not all logics have *analytic* (i.e., cut-free) proof systems, which are the *sine qua non* of proof search. A less obvious but equally bothersome limitation is that cut-free sequent proofs tend to contain large repeated sub-proofs.

To remedy these deficiencies, one can use the *calculus of structures*, a proof system that allows inferences anywhere inside a formula. This system can represent many more logics than the sequent calculus and can produce better (i.e., usually smaller) proofs because it can avoid sharing large subformulas. Nevertheless, because the rules of the calculus of structures have finer granularity than sequent rules, it has more non-determinism during search.

In this work, we show how to transplant the focusing result from the sequent calculus to the calculus of structures [19]. We thus improve the search capabilities of the calculus of structures, including the ability to go back and forth between focused sequent proofs and focused calculus of structures proofs, but we retain all the distinguishing features of the calculus of structures.

In particular, we preserve the ability to permute contractions below all other rules (first observed for the calculus of structures in [55], [31]). This permutation enables a two-stage normal form of proofs. The first stage contains only contractions, which increases the complexity of the formulas and is therefore a potential source of unbounded search; this phase needs to be recorded in order to reconstruct proofs by bounded search. The second stage that contains the remaining logical rules (except contraction) is strictly bounded and finite—hence decidable—and can be reconstructed if omitted from the proof object. Thus, we have the potential of obtaining very simple proof objects, recording only the first phase, for focused proofs; moreover, because of the bidirectional link, we can reconstruct focused sequent proofs from such proof objects.

Both the search and the representational aspects of focused calculus of structures proofs are being investigated in the Profound tool (see section 5.1).

6.2. Nested Sequents for Intuitionistic Modal Logics

Participant: Lutz Straßburger.

We show how the recent results for treating classical modal logics in the modal cube under S5 via nested sequents [32], [33] can be carried to intuitionistic modal logics. Thus, we present cut-free nested sequent systems for all intuitionistic modal logics in the modal cube up to IS5, and we show how this can be done in a modular way, i.e., to each of the axioms d, t, b, 4, and 5, we assign a set of inference rules, such that for each subset of d,t,b,4,5 the corresponding set of rules is sound and complete for the defined logic. This work has been presented in an invited talk at the IMLA workshop in Nancy [25].

6.3. The logic of nominal abstraction

Participant: Dale Miller.
In the paper [12], Gacek, Miller, and Nadathur have developed a strong logic that allows strong forms of induction and co-induction in the presence of the $\nabla$ quantifier. This quantifier was introduced in earlier work by Miller and Tiu [50] where it was shown to provide declarative and flexible operational semantic specifications for a number of systems such as the $\lambda$-calculus and the $\pi$-calculus. The paper introduces a generalization form of equality, called nominal abstraction, that permits natural specifications of predicates such as freshness. This paper provides the necessary meta theory (cut-elimination) for this new logic.

6.4. A framework for focusing and cut-elimination

**Participant:** Dale Miller.

Liang and Miller provide in [14] a general setting for specifying proofs in intuitionistic and classical logic. In this setting, it is possible to guarantee cut-elimination and initial elimination results simply by checking how certain simple side-conditions are used within a focused proof system. This setting treats both intuitionistic and classical logics as well allowing certain hybridizations of these two logics. This work helped lead the authors to finding a way to truly mix in one logic and one proof system both classical and intuitionistic logic, as described in [22].

6.5. Synthetic connectives and their proof system

**Participant:** Dale Miller.

In recent years, focused proof systems have been used to expand our understanding of how introduction rules and structural rule relate to each other. In these proof systems, inference rules and logical connectives are polarized as negative or positive in such a way that the invertible inference rules all belong to the negative polarity. Groups of negative connectives can then be grouped into one negative synthetic connective: similarly, positive connectives can be grouped into a positive synthetic connective. Such synthetic connectives admit cut-elimination. Remarkably, focused proof systems for classical and intuitionistic logics can be organized so that negative formulas are, in fact, treated linearly. That is, if weakening or contraction is applied to a formula, that formula is positive.

Focused proof systems can be used to design richly varying collections of synthetic connectives. These proof systems also provide for new means of describing parallelism within proofs and mixing computation and deduction. The ability to treat negative formulas linearly provides important information for the design of automated theorem provers. Synthetic connectives and their associated inference rules will also allow for the design of broad spectrum proof certificates that theorem provers will be able to print and simple proof checkers will be able to validate. Miller’s conference paper [20] develops this approach to proof certificates.

6.6. Automated reasoning and SMT solving

**Participants:** Mahfuza Farooque, Stéphane Lengrand.

Automated reasoning uses a broad range of techniques whose soundness and completeness relate to the existence of proofs. The research programme of the ANR PSI project at Parsifal is to build a finer-grained connection by specifying automated reasoning techniques as the step-by-step construction of proofs, as we know it from proof theory and logic programming. The goal is to do this in a unifying framework, namely proof-search in a classical polarized sequent calculus. One of the advantages of this is to combine those techniques more easily. Another one is to envisage extending those techniques.

For instance, the algorithm at the heart of SMT-solving (SAT-modulo-Theory) is DPLL(T), whose theory does not treat existential variables (although SMT-provers often implement incomplete ad hoc techniques for them). We have first encoded DPLL(T) as the step-by-step construction of a proof tree in a classical polarized sequent calculus extended with calls to a decision procedure for T (to be published). This proof-theoretic view now allows us to envisage how to extend the algorithm with existential variables.
Another range of techniques that we are addressing is the handling of equality (superposition / paramodulation calculi).

6.7. Towards a Stochastic Linear Logic for Biological Computation

Participants: Kaustuv Chaudhuri, Joëlle Despeyroux.

In previous work [35], Joëlle Despeyroux and Kaustuv Chaudhuri have given an encoding of the synchronous stochastic π-calculus in a hybrid extension of intuitionistic linear logic (called HyLL). Precisely, they have shown that focused partial sequent derivations in the encoding are in bijection with stochastic traces. The modal worlds are used to represent the rates of stochastic interactions, and the connectives of hybrid logic are used to represent the constraints in the stochastic transition rules.

One of the most successful applications of the stochastic π-calculus has been in representing signal transduction networks in cellular biology. An interesting application of this work would therefore be the direct representations of biological processes in HyLL, the original motivation for this line of investigation.

This year, we have worked on specifying some simple examples of regulatory gene networks, together with basic properties of them—such as stability or oscillation—in HyLL. This is ongoing work in collaboration with Gilles Bernot’s team at Nice-Sophia university.
6. New Results

6.1. Models of Programming

- P. Herms, together with C. Marché and B. Monate (CEA List), developed a certified VC generator, using Coq. The program for VC calculus and its specifications are both written in Coq, but the code is crafted so that it can be extracted automatically into a stand-alone executable. It is also designed in a way that allows the use of arbitrary first-order theorem provers to discharge the generated obligations [37]. This is a first step towards a certified VC generator for C programs annotated with ACSL [54].

- Until now, we only considered the proof of sequential programs. However the rely/guarantee approach gives a natural way to extend deductive verification to concurrent programs. During his internship supervised by C. Paulin, N. Gaspar explored this idea. He formalised in Coq the axiomatic semantics of a simple concurrent language using the rely-guarantee approach [43] and proposed a translation of this language into Why programs in order to automatically generate the proof obligations.

- W. Urribarri, together with C. Paulin, proposed an extension of the Why language with modules and functors together with a refinement calculus in order to organise large developments in a structured and abstract way. She built a prototype implementation of this calculus.

- D. Baelde, in cooperation with P. Courtieu (CNAM), D. Gross-Amblard (U. Bourgogne and Rennes), C. Paulin and X. Urbain proposed a formal proof of security for two watermarking algorithms. The proof uses a reduction of an arbitrary attack unmarking the data to the discovery of a secret key. It has been fully formalized in Coq using the ALEA library. This work has been presented at the conference TYPES 2011 and a paper is in preparation.

- Generating multimedia streams, such as in a netradio, is a task which is complex and difficult to adapt to every users’ needs. D. Baelde, in cooperation with R. Beauxis (Tulane University, LA, USA) and S. Mimram (CEA List) introduce a novel approach, based on a dedicated high-level functional programming language, called Liquidsoap, for generating, manipulating and broadcasting multimedia streams [20]. Unlike traditional approaches, which are based on configuration files or static graphical interfaces, it also allows the user to build complex and highly customized systems. This language is based on a model for streams and contains operators and constructions, which make it adapted to the generation of streams. The interpreter of the language also ensures many properties concerning the good execution of the stream generation.

6.2. Proofs of Imperative Programs

- The Why3 reimplementation of the Why platform, started in 2010, was publicly released in 2011 [35]. The main developers are A. Paskevich, J.-C. Filliâtre, F. Bobot, and C. Marché. The language of Why, both programming and annotation parts, was significantly extended: algebraic types, records, pattern matching, recursive logical definitions are now supported. These logical declarations are structured in modules (a.k.a. theories). The module language comes with an original mechanism for reusing theories in specialized contexts using partial instantiations. These new features have been presented at the first international workshop on intermediate verification languages (BOOGIE 2011) [21] and will be presented at VSTTE 2012 [69].
• A. Tafat and C. Marché used the Why3 system to perform a complete proof of the “Binary Heaps” challenge [41] from the VACID-0 international collection [76]. Solving this challenge is a case study for a general approach of abstract interfaces for C programs, currently under development by A. Tafat, based on initial ideas described together with S. Boulmé which were published in a 2011 special issue [29].

• In 2011, we set up a web gallery to publicly expose the programs that we specified and proved. This is the so-called ProVal collection of verified programs, and available at URL http://proval.lri.fr/gallery/index.en.html.

• K. Krishnamani and C. Marché proposed a technique for automatically generating loop invariants in C programs. It is based on the well-known predicate abstraction approach, which is adapted to take into account pre-existing specifications, and to proceed modularly, that is each function of a program is processed independently, with its own sets of predicates. The approach is also extended in order to generate universally quantified invariants [38]. The prototype is available as a Frama-C plugin at URL http://proval.lri.fr/agen.

• C. Dross, together with Y. Moy (Adacore) and J.-C. Filliâtre, addressed the problem of verifying programs involving containers. Containers such as lists, vectors, sets or maps are an attractive alternative to ad-hoc data structures based on pointers. C. Dross gave a definition of containers whose aim is to facilitate their use in certified software, using modern proof technology and novel specification languages. Correct usage of containers and user-provided correctness properties can be checked either by execution during testing or by formal proof with an automatic prover. It relies on a formal semantics for containers and an axiomatization of this semantics targeted at automatic provers. C. Dross proved in Coq that the formal semantics is consistent and that the axiomatization thereof is correct. This work was presented at TAP 2011 [26].

• Proving that pointer programs preserve data invariants, in a modular way, is known to be a challenge. R. Bardou and C. Marché designed a new approach based on advanced static typing systems (based on memory regions and permissions) to control memory updates and ownership of data [11]. A prototype implementation is built, called Capucine, available for download at http://romain.bardou.fr/capucine. To demonstrate the ability of this approach, the challenge “sparse arrays” of the VACID-0 benchmarks [76] has been certified [30].

• Separation logic has shown to be an elegant way to deal with programs which use data-structures with pointers. However it requires a specific logical language, provers, and specific reasoning techniques. In his PhD, F. Bobot introduced a technique to express ideas from separation logic in the traditional framework of deductive verification [12]. He proposed to derive “separation predicates” from user-supplied inductive definitions. These predicates come with suitable axiomatization, including frame rules, expressed in usual first-order logic. This translation takes special care to ensure the best use of automated theorem provers.

6.3. Automated Deduction

• In his thesis [14], S. Lescuyer formalized and designed purely reflexive tactics for automated deduction in Coq.

• É. Contejean, together with Pierre Courtieu, Julien Forest, Olivier Pons and Xavier Urbain (Cedric Laboratory, CNAM & ENSIIE) presented the last version of the rewriting toolkit CiME3 at RTA 2011 [25]. Amongst other original features, this version enjoys two kinds of engines: to handle and discover proofs of various properties of rewriting systems, and to generate Coq scripts from proof traces given in certification problem format in order to certify them with a skeptical proof assistant like Coq. Thus, these features open the way for using CiME3 to add automation to proofs of termination or confluence in a formal development in the Coq proof assistant.
• In their TACAS paper [24], S. Conchon, É. Contejean and M. Iguernelala present a modular extension of ground AC-completion for deciding formulas in the combination of the theory of equality with user-defined AC symbols, uninterpreted symbols and an arbitrary signature disjoint Shostak theory X.

• F. Bobot and A. Paskevich studied translation from a first-order logic with polymorphic types à la ML (which is the base logic of the Why platform and the Alt-Ergo theorem prover) to a many-sorted or one-sorted logic implemented in mainstream automated theorem provers. They devised a three-stage scheme where the last stage eliminates polymorphic types while adding the necessary “annotations” to preserve soundness, and the first two stages serve to protect certain terms so that they can keep their original types and unannotated form. Such protection allows to make use of provers’ built-in theories and operations. This work generalizes the previous study by S. Lescuyer and J.-F. Couchot [65] onto arbitrary monomorphic types, e.g. array types. It was presented at FroCoS 2011 [22] (see also an extended version with full proofs [42]). These results are part of F. Bobot’s PhD thesis [12].

6.4. Floating-Point and Numerical Programs

• T. Nguyen and C. Marché have worked on how to prove floating-point programs while taking into account architecture- and compiler-dependent features such as the use of the x87 stack in Intel micro-processors. This is done by analyzing the assembly code generated by the compiler [40], [28]

• S. Boldo and C. Marché published a survey article on the proofs of numerical C programs using both automatic provers and Coq [15].

• S. Boldo and T. Nguyen have worked on how to prove numerical programs on multiple architectures and compilers [17]. More precisely, it covers all the compiler choices about the use of extended registers, FMA, and reorganization of additions.

• S. Boldo and J.-M. Muller (CNRS, Arénaire, LIP, ÉNS Lyon) have worked on new floating-point algorithms for computing the exact and approximated errors of the FMA (fused multiply-and-add) [16].

• S. Boldo and G. Melquiond have developed in Coq a comprehensive formalization of floating-point arithmetic: core definitions, axiomatic and computational rounding operations, high-level properties [23]. It provides a framework for developers to formally certify numerical applications.

• G. Melquiond, in collaboration with F. de Dinechin (Arénaire, LIP, ÉNS Lyon) and C. Lauter (Intel Hillsboro), has improved the methodology for formally proving floating-point mathematical functions when their correctness depends on relative errors [19].

• S. Boldo, J.-C. Filliâtre and G. Melquiond, in collaboration with F. Clément (Estime, INRIA Paris-Rocquencourt) and M. Mayero (University Paris 13) have finished a full formal proof of a program solving a partial differential equation (the wave equation) using a finite difference scheme [36]. This proof includes both the mathematical convergence proof (method error) [57], a tricky floating-point proof [56] and proofs of the absence of runtime errors.

• C. Lelay, under the supervision of S. Boldo and G. Melquiond, has worked on differentiability in Coq. The goal was to prove the existence of a solution to the wave equation thanks to D’Alembert’s formula and to automatize the process as much as possible [44] [77].

• G. Melquiond, in collaboration with W. G. Nowak (Institute of Mathematics, Austria) and P. Zimmermann (Caramel, INRIA Nancy-Lorraine), has designed new methods for computing guaranteed enclosures of the Masser-Gramain constant, a two-dimensional analogue of the Euler-Mascheroni constant [86].

• G. Melquiond, in collaboration with J-M. Muller (CNRS, Arénaire, LIP, ÉNS Lyon) and E. Martin-Dorel (Arénaire, LIP, ÉNS Lyon), has worked on weakening the assumptions floating-point error-free transformations rely on [39].


6. New Results

6.1. White Noise-based Stochastic Calculus with respect to Multifractional Brownian Motion

Participants: Joachim Lebovits, Jacques Lévy Véhel.

The purpose of this work is to build a stochastic calculus with respect to (mBm) with a view to applications in finance and particularly to stochastic volatility models. We use an approach based on white noise theory.

6.1.1. White Noise-based Stochastic Calculus with respect to Multifractional Brownian Motion

The following results may be found in [28]. Integration with respect to mBm requires stochastic processes. Considering the probability space $(\mathbb{S}(\mathbb{R}), \mathbb{B}(\mathbb{S}(\mathbb{R})), \mu)$ where $\mu$ is probability measure given by Bôchner Minlos theorem, one can build to spaces, noted $(\mathbb{S})$ and $(\mathbb{S}^*)$ which will play an analogous role to the spaces $\mathbb{S}(\mathbb{R})$ and $\mathbb{S}'(\mathbb{R})$ for tempered distributions. We recall that $\mathbb{S}(\mathbb{R})$ is the Schwartz space of rapidly decreasing functions which are infinitely differentiable and $\mathbb{S}'(\mathbb{R})$ is the space of tempered distributions. Let us moreover note $(L^2)$ the space of random variables defined on the probability space $(\mathbb{S}(\mathbb{R}), \mathbb{B}(\mathbb{S}(\mathbb{R})), \mu)$ which admit a second order moment. The mBm $B^{(h)}$ has the following Wiener-Itô chaos decomposition in $(L^2)$:

\[
B^{(h)}(t) = \sum_{k=0}^{+\infty} <[0,t], \mathbb{M}_{h(t)}(e_k)>_{L^2(\mathbb{R})} = \sum_{k=0}^{+\infty} \left( \int_0^t \mathbb{M}_{h(t)}(e_k)(s) ds \right) <., e_k>
\]

where $(e_k)_{k \in \mathbb{N}}$ denotes the family of Hermite functions, defined for every integer $k$ in $\mathbb{N}$, by $e_k(x) := \pi^{-1/4}(2^k k!)^{-1/2} e^{-x^2/2} h_k(x)$ and where $(h_k)_{k \in \mathbb{N}}$ is the family of Hermite polynomial, defined for every integer $k$ in $\mathbb{N}$, by $h_k(x) := (-1)^k e^{x^2} \frac{d^k}{dx^k}(e^{-x^2})$. Note moreover that $\mathbb{M}_H$ is an operator from $\mathbb{S}(\mathbb{R})$ to $L^2(\mathbb{R})$ for every real $H$ in $(0, 1)$ and $<., e_k>$ is a centered random Gaussian variable with variance equal to 1 for all $k$ in $\mathbb{N}$. We can now define a process, noted $W^{(h)}$, from $\mathbb{R}$ to $(\mathbb{S}^*)$, which is the derivative of $B^{(h)}$ in sense of $(\mathbb{S}^*)$ by

\[
W^{(h)}(t) = \sum_{k=0}^{+\infty} \frac{d}{dt} \left( \int_0^t \mathbb{M}_{h(t)}(e_k)(s) ds \right) <., e_k>.
\]

Hence we define integral with respect to mBm of any process $\Phi : \mathbb{R} \to (\mathbb{S}^*)$ as being the element of $(\mathbb{S}^*)$ given by:

\[
\int_{\mathbb{R}} \Phi(s, \omega) dB^{(h)}(s) = \int_{\mathbb{R}} \Phi(s) \circ W^{(h)}(s) ds \ (\omega),
\]

where $\circ$ denotes the Wick product on $(\mathbb{S}^*)$. It is then possible to get Itô-Â’ formulas and Tanaka formula such as
\[
\int_0^T \frac{\partial f}{\partial x}(t, B^{(h)}(t)) dB^{(h)}(t) = f(T, B^{(h)}(T)) - f(0, 0) - \int_0^T \frac{\partial f}{\partial t}(t, B^{(h)}(t)) dt - \frac{1}{2} \int_0^T \left( \frac{d}{dt} [R_h(t, t)] \right) \frac{\partial^2 f}{\partial x^2}(t, B^{(h)}(t)) dt.
\] (7)

for functions with sub exponential growth and where the last equality holds in \( L^2 \).

Once this stochastic calculus with respect to \( mBm \) is defined, we can solve differential equations arising in mathematical finance.

6.1.2. Multifractional stochastic volatility

Multifractional stochastic volatility

The results of this part may be found in [6]. We assume that, under the risk-neutral measure, the forward price of a risky asset is the solution of the S.D.E.

\[
\begin{align*}
  dF_t &= F_t \sigma_t dW_t, \\
  d \ln (\sigma_t) &= \theta (\mu - \ln(\sigma_t)) dt + \gamma_\rho^h d^h B^h_t + \gamma_{\sigma} dW^\sigma_t, \quad \sigma_0 > 0, \quad \theta > 0,
\end{align*}
\] (8)

where \( W \) and \( W^\sigma \) are two standard Brownian motions and \( B^h \) is a multifractional Brownian motion independent of \( W \) and \( W^\sigma \) with functional parameter \( h \), which is assumed to be continuously differentiable.

We assume that \( W \) is decomposed into \( \rho dW^\sigma + \sqrt{1 - \rho^2} dW^F_t \), where \( W^F \) is a Brownian motion independent of \( W^\sigma \). Note that \( d^h B^h_t \) denotes differentiation in the sense of white Noise theory. The solution of the volatility process \( (\sigma_t)_{t \in [0,T]} \)

\[
\sigma_t = \exp \left( \ln(\sigma_0) e^{-\theta t} + \mu \left( 1 - e^{-\theta t} \right) + \gamma_\rho \int_0^t e^{\theta(s-t)} dW^\sigma_s + \gamma_h e^{-\theta t} I_t (B^h) \right),
\] (9)

where \( I_t (B^h) := e^{\theta t} B^h_t - \theta \int_0^t e^{\theta s} B^h_s ds \).

Since the solution the previous S.D.E. is not explicit for \( (F_t)_{t \in [0,T]} \) we use preconditioning and then cubature methods in order to get an approximation of it. This model allows to take into account the well-known "smile" effect of volatility, as well as its evolution at various maturities.

6.1.3. Approximation of \( mBm \) by \( fBms \)

In [18], we establish that a sequence of well-chosen lumped fractional Brownian motions converges in law to a multifractional Brownian motion. This allows to define stochastic integrals with respect to \( mBm \) by "transporting" corresponding stochastic integrals with respect to \( fBm \).

6.2. Sample paths properties of the set-indexed Lévy process

Participant: Erick Herbin.

In collaboration with Prof. Ely Merzbach (Bar Ilan University, Israel).

In [24], the class of set-indexed Lévy processes is considered using the stationarity property defined for the set-indexed fractional Brownian motion in [23]. Following Ivanoff-Merzbach’s definitions of an indexing collection \( \mathcal{A} \) and its extensions \( \mathcal{E}_0 = \{ U \upharpoonright V; \ U, V \in \mathcal{A} \} \) and

\[
\mathcal{E} = \left\{ U \upharpoonright \bigcup_{1 \leq i \leq n} V_i; \ n \in \mathbb{N}; \ U, V_1, \ldots, V_n \in \mathcal{A} \right\},
\]
a set-indexed process $X = \{X_U; U \in \mathcal{A}\}$ is called a set-indexed Lévy process if the following conditions hold

1. $X_{\emptyset'} = 0$ almost surely, where $\emptyset' = \bigcap_{U \in \mathcal{A}} U$.
2. the increments of $X$ are independent: for all pairwise disjoint $C_1, \cdots, C_n$ in $\mathcal{C}$, the random variables $\Delta X_{C_1}, \cdots, \Delta X_{C_n}$ are independent.
3. $X$ has $m$-stationary $\mathcal{C}_0$-increments, i.e. for all integer $n$, all $V \in \mathcal{A}$ and for all increasing sequences $(U_i)_i$ and $(A_i)_i$ in $\mathcal{A}$, we have

\[
\forall i, m(U_i \setminus V) = m(A_i) \Rightarrow (\Delta X_{U_1 \setminus V}, \cdots, \Delta X_{U_n \setminus V}) \overset{(d)}{=} (\Delta X_{A_1}, \cdots, \Delta X_{A_n})
\]

4. $X$ is continuous in probability.

On the contrary to previous works of Adler and Feigin (1984) on one hand, and Bass and Pyke (1984) on the other hand, the increment stationarity property allows to obtain explicit expressions for the finite-dimensional distributions of a set-indexed Lévy process. From these, we obtained a complete characterization in terms of Markov properties.

The question of continuity is more complex in the set-indexed setting than for real-parameter stochastic processes. For instance, the set-indexed Brownian motion can be not continuous for some indexing collection. We consider a weaker form of continuity, which studies the possibility of point jumps.

The point mass jump of a set-indexed function $x : \mathcal{A} \to \mathbb{R}$ at $t \in T$ is defined by

\[ J_t(x) = \lim_{n \to \infty} \Delta x_{C_n(t)}, \quad \text{where } C_n(t) = \bigcap_{C \in \mathcal{C}_n} C \]

and for each $n \geq 1$, $\mathcal{C}_n$ denotes the collection of subsets $U \setminus V$ with $U \in \mathcal{A}_n$ (a finite sub-semilattice which generates $\mathcal{A}$ as $n \to \infty$) and $V \in \mathcal{A}_n(u)$. A set-indexed function $x : \mathcal{A} \to \mathbb{R}$ is said pointwise-continuous if $J_t(x) = 0$, for all $t \in T$.

**Theorem** Let $\{X_U; U \in \mathcal{A}\}$ be a set-indexed Lévy process with Gaussian increments. Then for any $U_{\text{max}} \in \mathcal{A}$ such that $m(U_{\text{max}}) < +\infty$, the sample paths of $X$ are almost surely pointwise-continuous inside $U_{\text{max}}$, i.e.

\[ P(\forall t \in U_{\text{max}}, J_t(X) = 0) = 1. \]

In the general case, for all $\epsilon > 0$, For all $U \in \mathcal{A}$ with $U \subset U_{\text{max}}$, we define

\[ N_U(B) = \# \{ t \in U : J_t(X) \in B \}, \]

\[ X_U^B = \int_B x.N_U(dx), \]

for all $B \in \mathcal{B}_\epsilon$, the $\sigma$-field generated by the opened subsets of $\{ x \in \mathbb{R} : |x| > \epsilon \}$. The sample paths of the set-indexed Lévy processes can be derived from the following Lévy-Ito decomposition proved in [24].

**Theorem** Let $(\sigma, \gamma, \nu)$ the generating triplet of the SI Lévy process $X$.

Then $X$ can be decomposed as

\[ \forall \omega \in \Omega, \forall U \in \mathcal{A}, \quad X_U(\omega) = X_U^{(0)}(\omega) + X_U^{(1)}(\omega), \]
where

1. $X^{(0)} = \{X^{(0)}_U; U \in A\}$ is a set-indexed Lévy process with Gaussian increments, with generating triplet $(\sigma, \gamma, 0)$,

2. $X^{(1)} = \{X^{(1)}_U; U \in A\}$ is the set-indexed Lévy process with generating triplet $(0, 0, \sigma)$, defined for some $\Omega_1 \in \mathcal{F}$ with $P(\Omega_1) = 1$ by

   \[
   \forall \omega \in \Omega_1, \quad \forall U \in A, \quad X^{(1)}_U(\omega) = \int_{|x| > 1} x N_U(dx, \omega) + \lim_{\epsilon \downarrow 0} \int_{\epsilon < |x| \leq 1} x [N_U(dx, \omega) - m(U)] \nu(dx),
   \]

   where $N_U$ is defined in (13) and the last term of (14) converges uniformly in $U \subset U_{\text{max}}$ (for any given $U_{\text{max}} \in A$) as $\epsilon \downarrow 0$,

3. and the processes $X^{(0)}$ and $X^{(1)}$ are independent.

### 6.3. Hölder regularity of Set-Indexed processes

**Participants:** Erick Herbin, Alexandre Richard.

*In collaboration with Prof. Ely Merzbach (Bar Ilan University, Israel).*

In the set-indexed framework of Ivanoff and Merzbach ([54]), stochastic processes can be indexed not only by $\mathbb{R}$ but by a collection $A$ of subsets of a measure and metric space $(\mathcal{T}, d, m)$, with some assumptions on $A$. In [25], we introduce and study some assumptions on the metric indexing collection $(A, d_A)$ in order to obtain a Kolmogorov criterion for continuous modifications of SI stochastic processes. Under this assumption, the collection is totally bounded and a set-indexed process with good incremental moments will have a modification whose sample paths are almost surely Hölder continuous, for the distance $d_A$.

Once this condition is established, we investigate the definition of Hölder coefficients for SI processes. From the real-parameter case, the most straightforward are the local (and pointwise) Hölder exponents around $U_0 \in A$:

\[
\tilde{\alpha}_X(U_0) = \sup \left\{ \alpha : \limsup_{\rho \to 0} \sup_{U, V \in B_{d_A}(U_0, \rho)} \frac{|X_U - X_V|}{d_A(U, V)\alpha} < \infty \right\}.
\]

When the processes are Gaussian, a deterministic counterpart to this exponent is defined as it is in the real-parameter framework. For all $U_0 \in A$, we proved that almost surely, the random and the deterministic exponents are equal. Also, we proved that for the local exponents, this result holds almost surely, uniformly on $A$.

Given the particular structure of $A$, other coefficients of Hölder regularity were studied on $\mathcal{C}$:

\[
\mathcal{C} = \left\{ A \setminus \bigcup_{k=1}^{n} B_k : A, B_1, \cdots , B_n \in A, n \in \mathbb{N} \right\}.
\]

On specific subclasses $\mathcal{C}^l$ of $\mathcal{C}$ (satisfying $\bigcup_{l \geq 1} \mathcal{C}^l = \mathcal{C}$), the local (and pointwise) $\mathcal{C}^l$-Hölder exponents are defined:
\[ \tilde{\alpha}_{X,C}(U_0) = \sup \left\{ \alpha : \limsup_{\rho \to 0} \sup_{U \in B_{d,s}(U_0, \rho)} \frac{|\Delta X_{U \setminus V}|}{d_A(U,V)^{\alpha}} < \infty \right\}, \quad (14) \]

and this definition is proved to be independent of \( l \), leading to the definition of \( \tilde{\alpha}_{X,C}(U_0) \). It is compared to \( \tilde{\alpha}_{X}(U_0) \) and related to the Hölder exponent of the process projected on flows (a flow is a continuous increasing path in \( \mathcal{A} \)). This last technique permits to show that the pointwise Hölder exponent of the SIfBm is almost surely uniformly equal to \( H \), the Hurst parameter of the SIfBm. This completes some previous results on the multiparameter fractional Brownian motion.

The last exponent which is studied is the exponent of pointwise continuity:

\[ \alpha_{pc,X}(t) = \sup \left\{ \alpha : \limsup_{n \to \infty} \frac{|\Delta X_{C_n(t)}|}{m(C_n(t))^{\alpha}} < \infty \right\}, \quad (15) \]

for all \( t \in \mathcal{T} \), where \( C_n(t) \) is the smaller set of \( C_n \) containing \( t \). Almost sure results are also obtained in that case. For instance, the coefficient of pointwise continuity of a SI Brownian motion equals \( \frac{1}{2} \) a.s.

All these results are finally applied to the SIfBm and the SI Ornstein-Ühlenbeck process ([1]).

### 6.4. Stochastic 2-microlocal analysis

**Participants:** Erick Herbin, Paul Balança.

Stochastic 2-microlocal analysis has been introduced in [19] to study the local regularity of stochastic processes. If \( X = (X_t)_{t \in \mathbb{R}_+} \) is a stochastic process, then for all \( t_0 \in \mathbb{R}_+ \), a function \( s' \mapsto \sigma_{X,t_0}(s') \) called the 2-microlocal frontier is defined to characterize entirely the local regularity of \( X \) at \( t_0 \). In particular, for all \( s' \in \mathbb{R} \) such that \( \sigma_{X,t_0}(s') \in (0,1) \), it is defined as

\[ \sigma_{X,t_0}(s') = \sup \left\{ \sigma : \limsup_{n \to \infty} \sup_{u,v \in B(t_0,\rho)} \frac{|X_u - X_v|}{|u - v|^{\sigma}} < \infty \right\}. \]

The 2-microlocal frontier gives a more complete picture of the regularity than classical pointwise and local Hölder exponents, which are widely used in the literature. Furthermore, it is stable under the action of (pseudo-)differential operators.

[19] mainly focused on Gaussian processes, and in particular obtained a characterization of the regularity for Wiener integrals \( X_t = \int_0^t \eta_u dW_u \), with \( \eta \in L^2(\mathbb{R}) \).

Our main goal was therefore to extend this result to any stochastic integral

\[ X_t = \int_0^t H_u dM_u, \]

where \( M \) is a local martingale and \( H \) an adapted continuous process.

In fact, in [15], we first reduced this problem to the study of local martingales, and we have shown that almost surely for all \( t \in \mathbb{R}_+ \), the 2-microlocal frontier of a local martingale \( M \), with quadratic variation \( (M) \), satisfies
\[ \forall s' \geq -\alpha_{M,t}; \quad \sigma_{M,t}(s') = \Sigma_{M,t}(s') = \frac{1}{2} \Sigma_{M,t}(2s') , \]

where for any process \( X \), \( \Sigma_{X,t} \) denotes the pseudo 2-microlocal frontier which is characterized as following

\[ \forall s' \in \mathbb{R}; \quad \Sigma_{X,t}(s') = \sigma_{X,t}(s') \land (s' + p_{X,t}) \land 1 , \]

where \( p_{X,t} \) corresponds to

\[ p_{X,t} = \inf \left\{ n \geq 1 : X^{(n)}(t) \text{ exists and } X^{(n)}(t) \neq 0 \right\} , \]

with the usual convention \( \inf \{ \emptyset \} = +\infty \).

As the previous result is based on Dubins-Schwarz representation theorem, it can be easily extended to characterize the regularity of time-changed multifractional Brownian motions. In this case, we obtain a similar equation where \( \frac{1}{2} \) is replaced by \( H(t) \), the value of the Hurst function at \( t \).

Using this last equality, we can obtain the regularity of the stochastic integral \( X \) previously defined: almost surely for all \( t \in \mathbb{R}^+ \)

\[ \forall s' \geq -\alpha_{X,t}; \quad \sigma_{X,t}(s') = \Sigma_{X,t}(s') = \frac{1}{2} \sum_{\int_0^t H^2_M(u) dM_u,t} (2s') . \]

In the particular case of an integration with respect to a Brownian motion \( B \), the result can be simplified using the stability under differential operators: for almost all \( \omega \in \Omega \) and for all \( t \in \mathbb{R}^+ \), the 2-microlocal frontier satisfies

1. if \( H(t) \neq 0 \):

\[ \forall s' \in \mathbb{R}; \quad \sigma_{X,t}(s') = \sigma_{B,t}(s') = \left( \frac{1}{2} + s' \right) \land \frac{1}{2} , \]

2. if \( H(t) = 0 \):

\[ \forall s' \geq -\alpha_{X,t}; \quad \sigma_{X,t}(s') = \left( \frac{1}{2} + \frac{\Sigma_{H^2,t}(2s')}{2} \right) \land \frac{1}{2} , \]

unless \( H \) is locally equal to zero at \( t \), which induces in that case: \( \sigma_{X,t} = +\infty \).

Based on this last characterization, we were able to study the regularity of stochastic diffusions. In particular, we illustrated our purpose with the square of \( \delta \)-dimensional Bessel processes which verify the following equation

\[ Z_t = x + 2 \int_0^t \sqrt{Z_s} d\beta_s + \delta t . \]

6.5. Tempered multistable measures and processes

Participants: Jacques Lévy Véhel, Lining Liu.
This year, we concentrated on the following points:

- Define a new type of multistable processes called tempered multistable processes.
- Study the short time and long time behaviors of tempered multistable processes.
- Compare the multistable Lévy processes defined by finite-dimensional distributions (characteristic functions), Poisson representation and series representation.

The idea of the construction of tempered multistable measure and processes comes from the paper [63].

The interest of such processes is that they may be chosen to have moments of all orders. In addition, they are martingales. This will allow to construct stochastic (partial) differential equation driven by tempered multistable measures, which may be used to describe certain physical phenomena.

The characteristic function of a tempered multistable process $X(t)$ is

$$E(\exp iyX(t)) = \exp \left\{ \frac{1}{2} \int_0^t \Gamma(-\alpha(x)) \left[ \left( 1 - \frac{i y}{\theta} \right)^{\alpha(x)} + \left( 1 + \frac{i y}{\theta} \right)^{\alpha(x)} - 2 \right] \theta^{\alpha(x)} dx \right\}. $$

We have investigated the long time and short time behaviors this process:

Short time behavior:
Let $\alpha : \mathbb{R} \to [a, b] \subseteq (0, 2)$ be continuous. Let $u \in \mathbb{R}$ and suppose that as $v \to u$,

$$|\alpha(u) - \alpha(v)| = o \left( \frac{1}{|\log |u - v||} \right). \quad (16)$$

Then when $h \to 0$,

$$h^{-1/\alpha(t)}[X(t + hu) - X(t)] \to Y_{\alpha(t)}(u) \quad (17)$$

in finite-dimensional-distributions, where

$$Y_{\alpha(t)}(u) = \int 1_{[0,u]}(z) dM_{\alpha(t)}(z),$$

and $M_{\alpha(t)}$ is an $\alpha(t)$ stable measure. In an other word, $X(t) = M[0,t]$ is $1/\alpha(t)$-localisable at $t$ with local form $Y_{\alpha(t)}$.

Long time behavior:
Let $\alpha : \mathbb{R} \to [a, b] \subseteq (0, 2)$ be continuous and $\lim_{s \to \infty} \alpha(s) \to \alpha$. Then for $h \to \infty$

$$h^{-1/2}[X(t + hu) - X(t)] \to \Gamma(2 - \alpha) B(u) \quad (18)$$

in finite-dimensional-distributions, where $B$ is standard Brownian motion.

Let us now describe our work on the multistable Lévy motion. For $0 < a \leq b < 2$ and $\alpha : \mathbb{R} \to [a, b]$, the multistable Lévy motion $M_{\alpha}^c$ defined by finite-dimensional distributions (characteristics function) is the process such that

$$E(\exp i \sum_{j=1}^d \theta_j M_{\alpha, c}(t_j))) = \exp \left\{ - \int_0^t \left( \sum_{j=1}^d \theta_j 1_{[0,t_j]}(s) \right)^{\alpha(s)} ds \right\}; \quad (19)$$
There also exist a Poisson representation of multistable Lévy process $M_p$:

$$M_p(t) = \sum_{(X,Y) \in \Pi} C_{\alpha(X)} 1_{[0,t]}(X) Y^{<-1/\alpha(X)>}, \quad (20)$$

where $(X,Y)$ be the random point of the Poisson process $\Pi$, $t \geq 0$, $Y^{<-1/\alpha(X)>} = \text{sign}(Y)|Y|^{-1/\alpha(X)}$ and

$$C_{\alpha(X)} = \left( \frac{1}{\Gamma(1 - \alpha(X)) \cos (\frac{\pi}{2} \alpha(X))} \right)^{1/\alpha(X)}; \quad (21)$$

Finally, the series representation of multistable Lévy motion $M_s$ is

$$M_s(t) = \sum_{i=1}^{\infty} C_{\alpha(U_i)} \gamma_i^{1-1/\alpha(U_i)} 1_{(U_i \leq t)}, \quad (22)$$

where $\{\Gamma\}_{i \geq 1}$ is a sequence of arrival times of a Poisson process with unit arrival time, $\{U\}_{i \geq 1}$ is a sequence of i.i.d random variables with uniform distribution on $[0,t]$, $\{\gamma\}_{i \geq 1}$ is a sequence of i.i.d random variables with distribution $P(\gamma_i = 1) = P(\gamma_i = -1) = 1/2$. All three sequences $\{\Gamma\}_{i \geq 1}$, $\{U\}_{i \geq 1}$ and $\{\gamma\}_{i \geq 1}$ are independent, and

$$C_{\alpha(U_i)} = \left( \frac{1}{\Gamma(1 - \alpha(U_i)) \cos (\frac{\pi}{2} \alpha(U_i))} \right)^{1/\alpha(U_i)} \quad (23)$$

We have proved that these three definitions yield the same process in law.

### 6.6. Local strings and the CH set

**Participant:** Jacques Lévy Véhel.

*In collaboration with Prof. Franklin Mendivil (Acadia University, Canada).*

We have extended the definition of fractal strings originally proposed in [59] and modified in [37] to deal with the local behaviour of fractal sets. This allows to analyze the pointwise oscillatory properties of locally self-similar sets ([38]).

We have also analyzed in details the structure of a set build by "stacking" Cantor sets with continuously varying dimensions (see figure 4). The resulting set, called "Christiane’s hair" set or CH set, displays a number of interesting properties. Each "strand of hair" is a $C^\infty$ curve. Its Hausdorff dimension is 2. Furthermore, it is Minkowski measurable in dimension 2 with vanishing Minkowski content.

### 6.7. General models for drug concentration in multi-dosing administration

**Participants:** Lisandro Fermin, Jacques Lévy Véhel.

*In collaboration with P.E Lévy Véhel (University of Nice-Sophia-Antipolis and Banque Postale).*
In the past two years, we have developed models for investigating the probability distribution of drug concentration in the case of non-compliance. We have focused on two aspects of practical relevance: the variability of the concentration and the regularity of its probability distribution. In a first article [29], in a series of three, is considered the case of multi-intravenous dosing using the simplest possible law to model random drug intake, i.e. a homogeneous Poisson distribution. In a second article [13], we consider the more realistic multi-oral model, and deal with the complications brought by the first-order kinetics, which are essentially technical. Finally, in [12], we put ourselves in a powerful mathematical frame, known as Piecewise Deterministic Markov process (PDMP), that allows us to deal with general drug intake schedules, going beyond the homogeneous Poisson case. We use a PDMP to model the drug concentration in the case of multiple intravenous doses. In this particular model, we consider that the doses administration regimen is modeled by a non-homogeneous Poisson process whose jump rate is controlled by mean of a Markov chain. In this sense our PDMP model is a generalization to the continuos-models studied in [29]. In the following we detail our PDM model and the results obtained in the multi-IV case, see [12].

The model setting

Inspired by the PDMP model given in [47], [48], we consider a drug dosing stochastic regimen defined as follows.

Let us consider \((J_n)_{n\in\mathbb{N}}\) an irreducible Markov chain taking values in the state space \(K = \{1, \ldots, k\}\) with initial law \(\alpha_i = \mathbb{P}(J_0 = i)\) for all \(i \in K\) and transition probability matrix \(Q = (q_{ij})_{i,j \in K}\). We denote by \((T_n)_{n\in\mathbb{N}}\) the sequence of the random time doses and \((S_n)_{n\in\mathbb{N}}\) the time dose intervals; i.e. \(S_n = T_{n+1} - T_n\). We consider that the doses administration regimen is modeled by mean of the Markov process \((J_n)_{n\in\mathbb{N}}\) considering the following assumptions:

- The patient takes a dose \(D_{J_n} \in \{D_i, i \in K\}\) at the time \(T_n\), where the doses \(D_i\) are all different and different of zero.
- The time dose \(S_n\) is a random variable with exponential law of parameter \(\lambda_{J_n} \in \{\lambda_i, i \in K\}\), where the jump rate \(\lambda_i\) of state \(i\) is a positive constant.

We consider that these doses translate into immediate increases of the concentration by the value \(d_i = \frac{D_i}{V_d}\) if \(J_n = i\), where \(V_d\) is the apparent volume of distribution. After that, the effect of the dose taken at time \(T_n\) decreases exponentially fast with an exponential rate of elimination \(k_e\).

We define \((\nu_t)_{t \in \mathbb{R}}\) by \(\nu_t = \sum_{n>0} J_n \mathbb{I}_{[T_n, T_{n+1})}(t)\). We denote by \((C_t)_{t \in \mathbb{R}}\) the drug concentration stochastic process which take values on \(\mathbb{R}^*_+ = [0, \infty]\), we suppose that \(\mathbb{P}(C_0 = x) = 1\). Between the jumps, the dynamical evolution of the continuous time process \((C_t)\) is modeled by the flow \(\phi(t, x) = x \exp \{ -k_e t \} \).
Thus, the sample path of the stochastic process \((C_t)_{t \in \mathbb{R}_+}\) with values in \(\mathbb{R}_+^*\), starting from a fixed point \(x\) is given by

\[
C_t = xe^{-k_e t} + \sum_{i \geq 1} d_i e^{-k_e (t-T_i)} 1_{(t \geq T_i)}.
\] (24)

The process \((C_t, \nu_t)_{t \in \mathbb{R}_+}\) is a PDMP. From [49], we have that the infinitesimal generator \(\mathcal{U}\) of \((C_t, \nu_t)_{t \in \mathbb{R}_+}\) is given by

\[
\mathcal{U} f(x, i) = -k_e x \frac{d}{dx} f(x, i) + \lambda_i \sum_{j \in K} q_{ij} (f(x + d_j, j) - f(x, i)),
\] (25)

with \((x, i) \in E = \mathbb{R}_+^* \times K\) and \(f \in D(\mathcal{U})\) the set of measurable and differentiable on the first argument.

**The characteristic function of the concentration**

The characteristic function \(\varphi_\theta(t, x, i)\) of \(C_t\), given the starting point \((x, i)\), is the unique solution of the following system

\[
\begin{aligned}
\frac{\partial \varphi_\theta}{\partial t} (t, x, i) &= -k_e x \frac{\partial \varphi_\theta}{\partial x} (t, x, i) + \lambda_i \sum_{j \in K} q_{ij} \left( e^{i \theta d_j e^{-k_e t}} \varphi_\theta(t, x, j) - \varphi_\theta(t, x, i) \right), \\
\varphi_\theta(0, x, i) &= e^{i \theta x}.
\end{aligned}
\] (26)

**Variability of the concentration**

From (28) we have that the expectation \(m(t, x, i) = \mathbb{E}_{(x,i)}[C_t]\) of \(C_t\), given the starting point \((x, i)\), is given by

\[
m(t, x, i) = xe^{-k_e t} + \sum_{\nu, j \in K} \lambda_\nu q_{ij} d_j \int_0^t e^{-k_e (t-s)} P_{i\nu}(s) ds,
\] (27)

where \(P_{i\nu}(t) = \mathbb{P}(\nu_t = \nu|i_0 = i)\). The variance \(\text{Var}(t, i)\) of \(C_t\), given the initial state \(i\), is given by

\[
\text{Var}(t, i) = \sum_{\nu, j \in K} \lambda_\nu q_{ij} d_j^2 \int_0^t e^{-2k_e (t-s)} P_{i\nu}(s) ds - \left( \sum_{\nu, j \in K} \lambda_\nu q_{ij} d_j \int_0^t e^{-k_e (t-s)} P_{i\nu}(s) ds \right)^2 + 2 \sum_{\nu, j \in K} \sum_{j' \in K} \lambda_\nu q_{ij} d_j \lambda_{\nu j'} q_{i'j'} d_{j'} \int_0^t \int_0^{t-s} e^{-k_e (t-s)} P_{i\nu}(s) e^{-k_e (t-s-\tau)} P_{i\nu'}(\tau) d\tau ds.
\] (28)

**The distribution of limit concentration**

The characteristic function \(\varphi(\theta, i)\) of the limit concentration \(C\), given the starting state \(i\), satisfies

\[
-k_e \theta \frac{d}{d\theta} \varphi(\theta, i) + \sum_{j \in K} \lambda_j q_{ij} e^{i \theta d_j} \varphi(\theta, j) - \lambda_i \varphi(\theta, i) = 0.
\]

Thus, the random variables \(C(t)\) converge in distribution, when \(t\) tends to infinity, to a well-defined random variable \(C\) whose characteristic function is
\[ \varphi(\theta) = \sum_{j \in K} \varphi(\theta, j). \]

**Variability of the limit concentration**

We denote by \( m_i \) the mean of the limit concentration \( C \) in the state \( \nu = i \) and \( m = \sum_{i \in K} m_i \) the mean of \( C \) and \( \text{Var} \) its variance. Then,

\[
\begin{align*}
    m &= \frac{1}{\kappa} \sum_{i,j \in K} \pi_i \lambda_i q_{ij} d_j, \\
    m_i &= \frac{1}{\kappa} \sum_{j \in K} \pi_j \lambda_j q_{ji} d_i + \frac{1}{\kappa} \left( \sum_{j \in K} \lambda_j q_{ji} m_j - \lambda_i m_i \right), \\
    \text{Var} &= \frac{1}{2\kappa^2} \sum_{i,j \in K} \pi_i \lambda_i q_{ij} d_j^2 + \frac{1}{\kappa} \sum_{i,j \in K} \lambda_i q_{ij} d_j (m_i - \pi_i m) .
\end{align*}
\]

**Regularity of the limit concentration**

The characteristic function \( \varphi \) satisfies

\[
|\varphi(\theta)| \sim K |\theta|^{-\mu_{\text{max}}}, \quad \theta \to \infty,
\]

where \( K \) is a positive constant and \( \mu_{\text{max}} = \max_{i \in K} \frac{\lambda_i}{\kappa} \).

This result will allow us to describe in detail aspects of the limit distribution that are important for assessing the efficacy of therapy.

### 6.8. Complex systems design

**Participant:** Erick Herbin.

**In collaboration with Dassault Aviation, EADS, EDF.**

The preliminary design of complex systems can be described as an exploration process of a so-called design space, generated by the global parameters. An interactive exploration, with a decisional visualization goal, needs reduced-order models of the involved physical phenomena. We are convinced that the local regularity of phenomena is a relevant quantity to drive these approximated models. Roughly speaking, in order to be representative, a model needs more informations where the fluctuations are the more important (and consequently, where irregularity is the more important).

In collaboration with Dassault Aviation, EDF and EADS, we study how the local regularity can provide a good quantification of the concept of granularity of a model, in order to select the good level of fidelity adapted to a required precision.

Our works in that field can be expressed into:

- The definition and the study of stochastic partial differential equations driven by processes with prescribed regularity (that do not enter into the classical theory of stochastic integration).
- The study of the evolution of the local regularity inside stochastic partial differential equations (SPDE). Stochastic 2-microlocal analysis should provide informations about the local regularity of the solutions, in function of the coefficients of the equations. The knowledge of the fine behaviour of the solution of the SPDE will provide important informations in the view of numerical simulations.
6. New Results

6.1. Indistinguishability Proofs

Participants: Rohit Chadha, Vincent Cheval, Ştefan Ciobăcă, Hubert Comon-Lundh, Stéphanie Delaune, Steve Kremer.

Most existing results in verification of security protocols focus on trace properties such as secrecy or authentication. There are however several security properties that cannot be defined (or cannot be naturally defined) as trace properties and require the notion of indistinguishably. Typical examples are anonymity, privacy related properties or statements closer to security properties used in cryptography.

In the framework of the applied pi-calculus [54], as in similar languages based on equational logics, indistinguishability corresponds to a relation called trace equivalence. Roughly, two processes are trace equivalent when an observer cannot see any difference between the two processes.

Under some conditions, trace equivalence can be reduced to the problem of deciding symbolic equivalence, an equivalence relation introduced by M. Baudet [55]. However, the procedure proposed by Mathieu Baudet for deciding symbolic equivalence is complex and cannot be implemented in its current state. Moreover, this method can only deal with simple processes with trivial else branches and is restricted to the class of subterm-convergent equational theories. Unfortunately, this makes it unsuitable for some case studies of interest to the SECSI team, among which the FOO electronic voting protocol, and the electronic passport protocols.

In order to provide tool support to decide trace equivalence, Rohit Chadha, Stefan Ciobăcă, and Steve Kremer propose a procedure that can handle a large set of cryptographic primitives. The procedure has been implemented in a prototype tool and has been effectively tested on examples (e.g., the FOO e-voting protocol).

This paper is currently under submission.

Vincent Cheval, Hubert Comon-Lundh and Stéphanie Delaune have designed another procedure that allows one to check trace equivalence for a general class of processes [31]. In their class, they can model conditionals (with non-trivial else branches), private channels, and non-deterministic choice. The private authentication protocol and the various versions of the electronic passport protocol fall into their class.

6.2. Anonymous Credentials

Participants: Stéphanie Delaune, Malika Izabachène, Graham Steel.

Anonymous credentials plays an important role in non-interactive anonymous authentication: they allow a user to obtain certificates from organization and subsequently prove their possession in such a way that transactions of a same user remain unlinkable. In collaboration with Benoit Libert and Damien Vergnaud, Malika Izabachène present an anonymous credential scheme [39] in which a user can prove possession of appropriate attributes in an non-interactive fashion, by showing that these attributes satisfy a certain predicate (different type of predicates are handled).

Following this line of research on anonymous protocols, Stéphanie Delaune, Malika Izabachène and Graham Steel formalize unlinkability in the pi-calculus framework. They are exploring several scenarios in order to capture many adversarial strategies, especially in the context of low-cost devices, in which sensitive data are stored and identifier means are exchanged through public channels.

6.3. Security APIs

Participants: Stéphanie Delaune, Steve Kremer, Robert Künnemann, Graham Steel, Yusuke Kawamoto, Joe-Kai Tsay.
Security APIs allow untrusted code to access sensitive resources in a secure way. The idea is to design an interface between a trusted component, such as a smart card or cryptographic security module, and the untrusted outside world such that no matter what sequence of commands in the interface are called, and no matter what the parameters, certain good properties will continue to hold, e.g. the secret long term keys on the smartcard are never revealed. Designing such interfaces is very tricky, and several vulnerabilities in APIs in common use have come to light in recent years.

The members of the SECSI team have been studying the application of formal security analysis techniques to APIs, for the last few years. These APIs include industrial standards such as PKCS#11 and the Trusted Platform Module (TPM).

In [37], Delaune, Kremer and Steel present a Horn-clause-based framework for analyzing security protocols that use platform configuration registers (PCRs), which are registers for maintaining state inside the Trusted Platform Module (TPM). In their model, the PCR state space is unbounded, and experience shows that a naïve analysis using verification tools such as ProVerif or SPASS does not terminate. To address this, the authors extract a set of instances of the Horn clauses of the model, for which ProVerif does terminate on the chosen examples. The authors prove the soundness of this extraction process: no attacks are lost, that is, any query derivable in the more general set of clauses is also derivable from the extracted instances. The effectiveness of this framework is demonstrated in two case studies: a simplified version of Microsoft Bitlocker, and a digital envelope protocol that allows a user to choose whether to perform a decryption, or to verifiably renounce the ability to perform the decryption.

One of the reasons for the existence of security flaws that the members of the SECSI team identified when studying security APIs is the absence of definitions stating the expected security properties. In [40], Kremer, Steel and Warinschi propose a much-needed formal definition of security for cryptographic key management APIs. The advantages of this definition are that it is general, intuitive, and applicable to security proofs in both symbolic and computational models of cryptography. This definition relies on an idealized API which allows only the most essential functions for generating, exporting and importing keys, and takes into account dynamic corruption of keys. Based on this the authors can define the security of more expressive APIs which support richer functionality. They illustrate their approach by showing the security of APIs both in symbolic and computational models.

More recently, Kremer, Künnemann and Steel go even a step further in that direction and present the first key-management functionality in Canetti’s Universal Composability (UC) framework. It allows one to enforce a wide range of security policy and is highly extensible. The authors illustrate its use by proving an implementation of a Security API secure with respect to arbitrary key-usage operations and explore a proof technique that allows to store cryptographic keys externally, a novelty in the UC framework. This work is currently submitted.

In other recent work, in collaboration with Riccardo Focardi at the University of Venice, Kawamoto, Steel and Tsay have investigated the error behaviour of functions in the PKCS#11 API of various cryptographic devices including security tokens, electronic ID cards and Hardware Security Modules (HSMs). In certain circumstances attackers can take advantage of errors reported to make cryptanalytic attacks on functions in the API. Taking the example of the command used to import and encrypted key onto the device, they have discovered a number of so-called ‘error oracle attacks’ based on variations of well-known padding attacks due to Bleichenbacher and Vaudenay. This work has also recently been submitted. A number of vulnerability reports have been sent to manufacturers and national agencies.

### 6.4. Mobile Ad-Hoc Networks

**Participants:** Mathilde Arnaud, Morten Dahl, Stéphanie Delaune, Graham Steel.

Mobile ad hoc networks consist of mobile wireless devices which autonomously organize their communication infrastructure: each node provides the function of a router and relays packets on paths to other nodes. Finding these paths in an a priori unknown and constantly changing network topology is a crucial functionality of any ad hoc network. Specific protocols, called routing protocols, are designed to ensure this functionality known as
route discovery. Secure routing protocols use cryptographic mechanisms in order to prevent a malicious node from compromising the discovered route and they often perform some recursive tests on received messages.

Mathilde Arnaud, Véronique Cortier and Stéphanie Delaune provide \text{NPTIME} decision procedures for protocols with recursive tests and for a bounded number of sessions \cite{26}. They also revisit constraint system solving, providing a complete symbolic representation of the attacker knowledge.

In the context of vehicular ad-hoc networks, to improve road safety, a vehicle-to-vehicle communication platform is currently being developed by consortia of car manufacturers and legislators.

In \cite{35}, Morten Dahl, Stéphanie Delaune and Graham Steel propose a framework for formal analysis of privacy in location based services such as anonymous electronic toll collection. They give a formal definition of privacy, and apply it to the VPriv scheme for vehicular services. They analyse the resulting model using the ProVerif tool, concluding that the privacy property holds only if certain conditions are met by the implementation. Their analysis includes some novel features such as the formal modelling of privacy for a protocol that relies on interactive zero-knowledge proofs of knowledge and list permutations.

6.5. Composition Results

\textbf{Participants:} Céline Chevalier, Stéphanie Delaune, Steve Kremer.

Céline Chevalier, Stéphanie Delaune, and Steve Kremer investigate the composition of protocols that share a common weak secret \cite{32}. This situation arises when users employ the same password on different services. More precisely they study whether resistance against guessing attacks composes when a same password is used. More precisely, they present a transformation which maps a password protocol that is secure for a single protocol session (a decidable problem) to a protocol that is secure for an unbounded number of sessions. Their result provides an effective strategy to design secure password protocols: (i) design a protocol intended to be secure for one protocol session; (ii) apply the transformation and obtain a protocol which is secure for an unbounded number of sessions. This technique also applies to compose different password protocols allowing one to obtain both inter-protocol and inter-session composition.

6.6. Protecting Hypervisors from Denial of Service Attacks

\textbf{Participant:} Hedi Benzina.

Hedi Benzina showed that hypervisors can be protected from some denial of service attacks by allowing administrators to write security policies in a simple language \cite{41}. He implemented the RuleGen tool, which translates these policies into Orchids signatures.

6.7. Soundness Results: Some Limitations

\textbf{Participant:} Hubert Comon-Lundh.

Soundness results aim at bridging the gap between computational and symbolic security; they show that some symbolic model, in which messages are terms and the attacker is a formal process, faithfully abstracts the computational model, in which messages are bitstrings and the attacker is any probabilistic polynomial time Turing machine. Such results allow one to derive strong security guarantees, while reasoning at an abstract level. They have been developed for several cryptographic primitives (e.g. symmetric and asymmetric encryption, signatures, hash) and security properties.

These results however suffer from some severe limitations, as Hubert Comon-Lundh and Véronique Cortier demonstrate \cite{34}, focusing on symmetric encryption.

6.8. Model-Checking Reactive Probabilistic Systems

\textbf{Participant:} Rohit Chadha.
Rohit Chadha along with A. Prasad Sistla and Mahesh Viswanathan continued their study on reactive probabilistic systems modeled as Probabilistic Büchi Automata (PBA) in [30]. Reactive probabilistic systems are probabilistic non-deterministic systems in which the nondeterminism is resolved by a external environment which is oblivious of the “current” state of the system. This paper investigates the power of PBA when the threshold probability of acceptance is non-extremal, i.e., is a value strictly between 0 and 1. Many practical randomized algorithms are designed to work under non-extremal threshold probabilities and thus it is important to study power of PBAs for such cases. The paper presents a number of surprising expressiveness and decidability results for PBAs when the threshold probability is non-extremal. Some of these results sharply contrast with the results for extremal threshold probabilities. The paper also presents results for Hierarchical PBAs and for an interesting subclass of them called simple PBAs.

Rohit Chadha along with V. Korthikranthi, M. Viswanathan, G. Agha and Y. Kwon also study reactive probabilistic systems in [28]. In [28], reactive probabilistic systems are viewed as transformers of probability distributions, giving rise to a labeled transition system over the probability distributions over the states of the system. Thus, a reactive probabilistic system can be seen as defining a set of executions where each execution is a sequence of probability distributions. Reasoning about sequences of distributions allows one to express properties not expressible in standard probabilistic logics like PCTL; examples include expressing bounds on transient rewards and expected values of random variables, as well as comparing the probability of being in one set of states at a given time with another set of states. With respect to such a semantics, the model-checking problem is undecidable. In this paper, the authors identify a special class of systems called semi-regular Markov Decision Processes that have a unique non-empty, compact, invariant set of distributions, for which they show that checking any $\omega$-regular property is decidable. Their decidability result also implies that for semi-regular probabilistic finite automata with isolated cut-points, the emptiness problem is decidable.

6.9. Continuous Random Variables

Participant: Jean Goubault-Larrecq.

Continuing work on probabilistic and non-deterministic choice in a domain-theoretic setting, Jean Goubault-Larrecq and Daniele Varacca (PPS, University Paris 7) proposed a new monad for probabilistic choice, that of continuous random variables [38]. The usual Jones-Plotkin monad of continuous valuations, although simple enough, suffers from the defect that no category of continuous domains is known that would be both Cartesian-closed (i.e., would allow one to interpret functions) and stable under the Jones-Plotkin monad. Jean Goubault-Larrecq and Daniele Varacca managed to show that a related monad, that of continuous random variables, inspired from the notion of a random variable in probability theory, did not suffer from this defect: the category of bc-domains is indeed both Cartesian-closed and stable under this monad. Moreover, the authors showed that using one or the other monad gave semantics to higher-order probabilistic programs that were indistinguishable at ground types. Finally, they used this to solve an open problem by Escardò, namely that observational equivalence of probabilistic higher-order programs is recursively enumerable.

6.10. Choquet-Kendall-Matheron Theorems

Participant: Jean Goubault-Larrecq.

One of the results obtained by Jean-Goubault-Larrecq in his theory of semantics for mixed non-deterministic and probabilistic choice [60] is that there is a one-to-one correspondence between continuous credibilities over some (state) space $X$ and certain compact subsets of the space of all continuous valuations over $X$, under mild assumptions on $X$. Similar theorems were produced by Choquet in the 1950s, refined by Kendall, then by Matheron in the 1970s, with applications in random set theory, among others.

Klaus Keimel and Jean Goubault-Larrecq produced an extremely simple proof of this fact [22], based on a simple special case of Groemer’s integral theorem. This proof also produces a much more general result than what was known earlier, as it does not assume that $X$ is second-countable or Hausdorff, and only local compactness.
A domain-theoretic view is that this is a representation theorem for mixed demonic choice and probabilistic choice; the angelic and erratic cases are also covered by Goubault-Larrecq and Keimel. These results had been presented at Dagstuhl Seminar 10232, June 2010.

### 6.11. Full Abstraction for Call-by-Value Programs with Choice

**Participant:** Jean Goubault-Larrecq.

Consider a programming language, with both an operational semantics, stating how one can implement a machine for this language, and a denotational semantics, which states what programs compute (not how). A classical question in programming language semantics is whether equality of denotations (from denotational semantics) coincides with contextual equivalence (from operational semantics). This is called full abstraction.

This question was first formulated for PCF by G. Plotkin in 1977, who showed that PCF was not fully abstract, although PCF plus a form of parallel or was. PCF is a simply-typed higher-order language, which one could see as a simple variant of the ML language without mutable state.

Jean Goubault-Larrecq examined the question for variants of PCF with various forms of non-deterministic and probabilistic choice. The latter are modeled denotationally by using his theory of previsions [61]. The most startling result is that the call-by-value variant of PCF with only angelic non-determinism is fully abstract, without the need for parallel or. Jean Goubault-Larrecq also showed that call-by-value PCF with angelic non-determinism and probabilistic choice is not fully abstract, but that this language plus so-called statistical test primitives is fully abstract. These results were presented at the Domains X Workshop, Swansea, Wales, UK, September 2011.
SELECT Project-Team

6. New Results

6.1. Model selection in Regression and Classification

Participants: Gilles Celeux, Mohammed El Anbari, Clément Levrard, Robin Genuer, Erwan Le Pennec, Lucie Montuelle, Pascal Massart, Caroline Meynet, Jean-Michel Poggi.

Erwan Le Pennec continues his work with Serge Cohen (IPANEMA Soleil) on hyperspectral image segmentation based on a spatialized Gaussian Mixture Model. They derive, and implement within MIXMOD, an efficient minimization algorithm combining EM algorithm, dynamic programming and model selection[37]. They have applied this technique to analyze ancient material[9] This scheme is supported by a theoretical work on conditional density estimation[40]. In the framework of her PhD, Lucie Montuelle has studied some extension to this model to spatial varying logistic weights.

In collaboration with Marie-Laure Martin-Magniette (URGV et UMR AgroParisTech/INRA MIA 518) and Cathy Maugis (INSA Toulouse) has extended their variable selection procedure for model-based clustering and supervised classification to deal with high dimensional data sets with a backward selection procedure which is more efficient that the previous forward selection procedure in this context. [17]. Moreover they have shown the advantage of the model-based approach over a geometrical approach to select variable for clustering [13]. These variable selection procedures are in particular used for genomics applications which is the result of a collaboration with researchers of of URGV (Evry Genopole).

Caroline Meynet provided an $\ell_1$-oracle inequality satisfied by the Lasso estimator with the Kullback-Leibler loss in the framework of a finite mixture of Gaussian regressions model for high-dimensional heterogeneous data where the number of covariates may be much larger than the sample size. In particular, she has given a condition on the regularization parameter of the Lasso to obtain such an oracle inequality. This oracle inequality extends the $\ell_1$-oracle inequality established by Massart and Meynet [16] in the homogeneous Gaussian linear regression case. It is deduced from a finite mixture Gaussian regression model selection theorem for $\ell_1$-penalized maximum likelihood conditional density estimation, which is inspired from Vapnik’s method of structural risk minimization and from the theory on model selection for maximum likelihood estimators developed by Massart.

From an practical point of view, Caroline Meynet has introduced a procedure to select variables in model-based clustering in a high-dimensional context. In order to tackle with the problem of high-dimension, she has proposed to first use the Lasso in order to select different sets of variables and then estimate the density by a standard EM algorithm by reducing the inference to the linear space of the selected variables by the Lasso. Numerical experiments show that this method can outperform direct estimation by the Lasso.

In collaboration with Professor Abdallah Mkhadri (University of Marrakesh, Marocco), Gilles Celeux supervised the thesis of Mohammed El Anbari which concern regularisation methods in linear regression. In collaboration with Professor Abdallah Mkhadri (University of Marrakesh, Marocco), Mohammed El Anbari proposed a method to simultaneously select variables and favor a grouping effect where strongly correlated predictors tend to be in or out of the model together. Numerical experiments showed that their method can be preferred to Elastic-Net when the number of variables is less or equal to the sample size and remain competitive otherwise. Moreover, they have proposed AdaGril an extension of the the adaptive Elastic Net which incorporates information redundancy among correlated variables for model selection and estimation. Under weak conditions, They have established an oracle property of AdaGril. Numerical experiments show in some cases of AdaGril outperforms its competitors.

In collaboration with Jean-Michel Marin (Université de Montpellier) and Christian P. Robert (CEREMADE, Université Paris Dauphine) Gilles Celeux and Mohammed El Anbari highlight the interest of Bayesian regularization methods, using hierarchical non informative priors, compared with standard regularization methods in a poorly informative context through numerical experiments [47].
Clément Levrard worked on the obtention of fast rates of convergence for vector quantization. Using theoretical analogies between quantization seen as an unsupervised learning problem and the one of supervised learning by empirical contrast minimization, he has obtained a logarithmic improvement on the previously obtained bound. He has been furthermore able to define intelligible "margin type" condition under which fast rates can be obtained.

Since September 2008, Pascal Massart is the cosupervisor with Frédéric Chazal (GEOMETRICA) of the thesis of Claire Caillerie (GEOMETRICA). The project intends to explore and to develop new researches at the crossing of information geometry, computational geometry and statistics.

6.2. Statistical learning methodology and theory

Participants: Gilles Celeux, Christine Keribin, Erwan Le Pennec, Pascal Massart, Lucie Montuelle, Jean-Michel Poggi.

Unsupervised segmentation is an issue similar to unsupervised classification with an added spatial aspect. Functional data is acquired on points in a spatial domain and the goal is to segment the domain in homogeneous domain. The range of applications includes hyperspectral images in conservation sciences, fMRI data and all spatialized functional data. Erwan Le Pennec and Lucie Montuelle are focusing on the questions of the way to handle the spatial component from both the theoretical and the practical point of views as well as the choice of the number of clusters. Furthermore, as functional data require heavy computation, they are required to propose numerically efficient algorithms.

Gilles Celeux, Christine Keribin and the Ph.D. student Vincent Brault continue their work on the Latent Block Model. They have proposed an efficient algorithm coupling a Stochastic version of the EM algorithm including a Gibbs sampling step and the Variational EM algorithm. This SEM-VEM algorithm is insensitive to its initial position. On the other hand they got a closed formed expression of the Integrated Completed Likelihood for binary tables which allows for a reliable model selection criterion avoiding asymptotic approximation. Moreover, Christine Keribin derived sufficient conditions ensuring the identifiability of the Latent Block Model.

6.3. Reliability and Computer Experiments

Participants: Yves Auffray, Gilles Celeux, Rémi Fouchereau, Shuai Fu, Pascal Massart.

In the computer experiments field, the goal is to approximate an expensive black box function from a limited number of evaluations. The choice of these evaluations i.e. the choice of a design of (computer) experiments is a major issue.

This year Yves Auffray and Pierre Barbillon, in collaboration with Jean-Michel Marin (Université de Montpellier) have considered estimating the probability of rare events in the context of computer experiments. These rare events depend on the output of a physical model with random input variables. Since the model is only known through an expensive black box function, a crude Monte Carlo estimator does not perform well. Two strategies have been developed to cope with this difficulty: a Bayesian estimate and an importance sampling method. Both methods relies on Kriging metamodeling. They are able to achieve sharp upper confidence bounds on the rare event probability. These methods have been applied to a toy example and a real case study which consists of finding an upper bound of the probability that the trajectory of an airborne load collides the aircraft that has released it.

Following the previous work of the first year, Shuai Fu, under the direction of Gilles Celeux, focus on the design of experiments and its validation, which has become the main issues of the thesis. It leads both to theoretical and computational developments. An original DAC criterion has been proposed and leads to a Bayesian procedure of DAC-test to measure the quality of a design. For improving the design of experiments, an adaptive kriging procedure well adapted to the specific problem has been proposed. However, the algorithms require a too important computation time which should be reduced in future work.
In the framework of a CIFRE convention with Snecma-SAFRAN Rémy Fouchereau has started a thesis on the modeling of fatigue damage for Inco718 supervised by Gilles Celeux. Inco718 is a Zinc-based alloy. To determine its minimum lifetime, a lot of stress tests are made. The lloay lifetimes are reported as function of the stress. The aim of this work is to analyse the resulting curves. A mixture model with a lognormal component and a sum of two lognormals components is considered. Since the sum of two or more lognormal distribution is not closed form. Inference on this model needs Monte Carlo integration within the EM algorithm. Despite some unstability for small sample sizes, this model show encouraging and easily interpretable results.

6.4. Statistical analysis of genomic data

Participants: Gilles Celeux, Andrea Rau.

Andrea Rau and Gilles Celeux, in collaboration with Marie-Laure Martin-Magniette (URGV and UMR AgroParisTech/INRA MIA 518) and Cathy Maugis-Rabusseau (IMT/INSA Toulouse) have developed a method to cluster digital gene expression observations from high-throughput (HTS) data using Poisson mixture models [44]. The proposed model has the advantage of accounting for the particularities of HTS data and providing straightforward procedures for parameter estimation and model selection. A series of simulation experiments was done to compare the performance of the proposed model to that of previously proposed clustering methods for similar sequence-based data, and the performance of the proposed approach was examined on two real high-throughput sequencing data sets. The R package HTSCluster is used to implement the proposed Poisson mixture model has been made freely available on CRAN.

6.5. Curves classification, denoising and forecasting

Participant: Jean-Michel Poggi.

In collaboration with Farouk Mhamdi and Meriem Jaidane (ENIT, Tunis, Tunisia), Jean-Michel Poggi proposed, in [18], a method for trend extraction from seasonal time series through the Empirical Mode Decomposition (EMD). Experimental comparison of trend extraction based on EMD, X11, X12 and Hodrick Prescott filter are conducted. First results show the eligibility of the blind EMD trend extraction method. Tunisian real peak load is also used to illustrate the extraction of the intrinsic trend.

In collaboration with Mina Aminghafari (Amirkabir University, Teheran), Jean-Michel Poggi made uses of wavelets in a statistical forecasting purpose for time series. Recent approaches involve wavelet decompositions in order to handle non stationary time series. They study and extended an approach proposed by Renaud et al., to estimate the prediction equation by direct regression of the process on the Haar non-decimated wavelet coefficients depending on its past values. The new variants are used first for stationary data and after for stationary data contaminated by a deterministic trend [3].

Jean-Michel Poggi was the supervisor (with A. Antoniadis) of the PhD Thesis of Jairo Cugliari-Duhalde which takes place in a CIFRE convention with EDF. It is strongly related to the use of wavelets together with curves clustering in order to perform accurate load consumption forecasting. The thesis develops methodological and applied aspects linked to the electrical context as well as theoretical ones by introducing exogeneous variables in the context of nonparametric forecasting time series (see [27] and [45]).

6.6. Neuroimaging, Statistical analysis of fMRI data

Participants: Gilles Celeux, Christine Keribin.

This research takes place as part of a collaboration with Neurospin on brain functional Magnetic Resonance Imaging (fMRI) data. (http://www.math.u-psud.fr/select/reunions/neurospin/Welcome.html). This year it concerned essentially regularisation in a supervised clustering methodology that includes spatial information in the prediction framework, and yields clustered weighted maps.
6. New Results

6.1. Point counting

In joint work with Pierrick Gaudry (CARAMEL) and David Kohel (Marseille), B. Smith developed an accelerated Schoof-type point counting algorithm for genus 2 curves with efficiently computable real multiplication endomorphisms. This project has made the computation of cryptographic-sized group orders practical for curves of genus 2 over prime finite fields. Going way beyond the current cryptographic range, the algorithm has been used to compute the group order of a 1024-bit Jacobian (smashing the previous 256-bit record of Gaudry and Schost). The article describing this algorithm has been awarded the Best Paper prize at ASIACRYPT 2011 [26], and an extended version has been invited for submission to Journal of Cryptology (the leading journal in the field).

6.2. Complex multiplication

F. Morain has been investigating new invariants for building class polynomials with small coefficients. This is still work in progress, though advertised in some talks of his.

6.3. Steganography

D. Augot, M. Barbier and Caroline Fontaine randomized the bounded syndrome coding problem on wet paper—an important embedding problem in steganography—such that this problem always has a solution [24]. This randomization is inspired the Courtois–Finiasz–Sendrier signature scheme, and shows nice results for linear perfect codes. In the special case of binary Hamming codes, this new method reaches exactly the necessary and sufficient bounds to ensure the embedding. The previous bounds were introduced by Carlos Munuera and M. Barbier [19]. These bounds depend on the dual distance of the code used. Thanks to the generalized Hamming weight, they proved that codes with low MDS rank are better in this context. Since the nature of their results are combinatorial, the authors generalized a bound for systematic non linear codes and showed that the non-linear systematic codes could be good candidates, as shown by the example of the Nadler code.

6.4. Homomorphic encryption

D. Augot, in collaboration with L. Perret from Salsa team, and Bochum Universität [22], designed a “secret-key” homomorphic encryption scheme, which is much more efficient than the public-key ones. It is based on \(q\)-ary Reed-Muller codes (or multi-variate evaluation-interpolation schemes). The main drawback is a severe restriction on the number of uses of a given secret key, but the ease of decrypting leads to think that the scheme can reencrypt its keys, enabling its reuse.

6.5. List decoding

D. Augot, M. Barbier and A. Couvreur wrote on how to decode binary Goppa codes. Augot, Barbier, and Couvreur presented a simple way, with a clean study of the complexity [23]. Using this list decoding algorithm, Barbier and Paulo Barreto proposed a key reduction for the McEliece cryptosystem [25]. The list decoding algorithm above allowed them to add more errors during the McEliece encryption step, making decoding attacks more difficult. At the same complexity of these attacks, using the list decoding algorithm decreases the public key size, which is the main drawback of this cryptosystem.
6.6. Explicit isogeny constructions

B. Smith constructed six infinite series of families of pairs of algebraic curves of arbitrarily high genus [20], defined over number fields, together with an explicit isogeny between the Jacobians of the curves splitting multiplication by 2, 3, or 4.

6.7. Quasi-cyclic codes

M. Barbier, Christophe Chabot and G. Quintin exhibited a bijective correspondence between the $\ell$-quasi-cyclic codes over $\mathbb{F}_q$ of length $m\ell$ and the set of ideals of $M_\ell(\mathbb{F}_q)[X]/(X^m - 1)$ [29]. They proposed also two new classes called the quasi-BCH and quasi-evaluation codes. For the first one, they introduced a unambiguous decoding algorithm, and thanks to the second one they designed 49 new codes over $\mathbb{F}_4$ which have a bigger minimum distance than previously known codes.

6.8. Root-finding over Galois rings

Jérémy Berthomieu, Grégoire Lecerf and G. Quintin presented a new algorithm to find all the roots of a given polynomial with coefficients in a Galois ring [30]. It has been used to study the behavior of the Sudan algorithm for Reed-Solomon codes over Galois rings. The algorithm has been adapted to work over rings of power series in several variables. It was implemented in the Quintix package of Mathemagix.
6. New Results

6.1. Autonomous Computing

Participants: Cécile Germain-Renaud, Michèle Sebag, Balázs Kégl, Yusik Kim, Julien Nauroy, Dawei Feng.

Within the classical objectives of Autonomics (self-*), two transversal lines of research have emerged.

Coping with non-stationarity  Most existing work on modeling the dynamics of grid behavior assumes a steady-state system and concludes to some form of long-range dependence (slowly decaying correlation) in the associated time-series. But the physical (economic and sociologic) processes governing the grid behavior dispel the stationarity hypothesis. [ 68 ] proposes a categorization of the methods integrate non-stationarity into grid modeling. [ 9 ] considers a specific class of models: a sequence of stationary processes separated by breakpoints. The model selection question is now defined as identifying the breakpoints and fitting the processes in each segment, together with a validation methodology that empirically addresses the current lack of theoretical results concerning the quality of the estimated model parameters. Even when stationarity is acceptable, the markovian assumption might be too bold. [ 54 ] integrate Echo State Network-based regression into a reinforcement learning in continuous state space for fitting the Q function, with application to reactive grid scheduling.

Traces mining  In order for an autonomic system to continuously infer knowledge from its monitoring (the so-called MAPE-K, Monitor-Analyze-Plan-Execute-Knowledge) loop, heterogeneous sources of data have to be integrated. [ 37 ] exemplifies two use cases of the Grid Observatory data on evaluating the performance of the major EGI scheduler, and blackhole detection.

The Green Computing Observatory [ 38 ] data include the detailed monitoring of the processors and motherboards, as well as the global site information, such as overall consumption and overall cooling. The data schema for publication is grounded in an ontology of measurements developed in collaboration with the MIS (Modélisation, Information et Systèmes) laboratory of University Picardie Jules Verne.

[ 42 ] proposes a new approach for analyzing behavioral traces: as most of them are indeed text documents, state of the art techniques in text mining, including Latent Dirichlet Allocation, can be exploited. The advantages are twofold: providing some level of explanation inferred from the data; and a relatively scalable way to capture the temporal variability of the behavior of interest, while retaining the full dimensionality of the problem at hand. A promising perspective for combining this approach and inferred segmentation has been identified and is currently explored.

6.2. Complex Systems

Participants: Jamal Atif, Nicolas Bredèche, Matthias Brendel, Cyril Furtlehner, Philippe Caillou, Jean-Marc Montanier, Hélène Paugam-Moisy, Marc Schoenauer, Michèle Sebag.

Evolutionary AI Planning: Divide And Evolve (DAE) DAE solves AI-planning problems by using an Evolutionary Algorithm to sequentially divide them into hopefully simpler problems that are handled by some embedded “classical” planner. Within the ANR project DESCARWIN, work has continued in collaboration with Thalès Research & Technology and ONERA Toulouse. A large part of the work this year has been devoted to writing a brand new version of the DAE software, facing difficulties of parallelization [ 90 ]. The resulting program entered the 7th International Planning Competition (IPC 2011) at the 21st International Conference on Planning and Scheduling (ICAPS 2011) and won the Gold Medal in the Temporal Track. Note that the Silver Medal was won by Vincent Vidal,
also member of the DESCARWIN team, using his planner YAHSP2 – the one that won the Gold Medal while embedded in DAE, thus demonstrating one more the added value of the DAE approach. Meanwhile, because DAE has many parameters (like most Evolutionary Algorithms), parameter tuning within DAE remains a difficult task, and an original approach has been proposed to learn the parameters based on some instance features \cite{49}, \cite{50}, \cite{51}. Note that this method is however relevant of the “Crossing the Chasm” SIG (see Section 6.3), as it can be applied to any optimization algorithm that handles several instances of the same class.

Distributed Autonomous Robotics Resuming work done in 2010, we investigated further the issue of robotic swarm control whenever the environment is partially or completely unknown. This research is at the cross-road of Evolutionary Computation, Machine Learning and Robotics, and a light influence from Evolutionary Ecology, but with a strong focus on engineering (ie. the goal remains to design algorithms). The topic we are interested in is the design of environment-driven self-adaptive distributed algorithms to enable survival at the level of a population of independent robotic units. The population is limited in size, and hardware implementation within real robots has already been achieved \cite{7}. We have also focused our attention on specific aspects of swarm evolutionary dynamics under specific constraints, including the evolution of cooperative and/or altruistic behaviours \cite{53}, \cite{52}. This research yielded interesting results, such as the emergence of altruistic behavior under simple, but specific, algorithmic constraint, as well as tuning mechanism to control the level of altruistic behavior in a population of robots. Perspectives of this work is currently under investigation.

The work done in 2010 about the division of labor among asynchronous and decentralized agents, where each agent is modelled from the competition between two spiking neurons, was further analyzed within a spatio-temporal (simulated) frame. The phase transitions between the asynchronous, the aperiodic and periodic synchronous regimes (depending on the sociability and excitability of the agents) was confirmed, with some counter-intuitive results about the overall merits and efficiency of synchronous behaviors \cite{23}.

We have also explored objective-driven online learning within real robotic hardware, both for single robot online behavior learning \cite{78} as well as small group of robots for pattern formation learning \cite{43}. Our activity in Evolutionary Robotics has also been strengthened by the publication of book which gather several contributions from major actors in the field \cite{75}, including an introduction paper on current trends and challenges in this domain \cite{76}.

From a slightly different perspective, our work on evolving generative and developmental representations has been continued, with an extensive study of robustness within developmental systems \cite{8} and an investigation of the temporal dynamics at work within genetic regulatory networks for design \cite{32}. While not stricly related to robotics, these contributions share the distributed nature of computation and ultimately aim at providing an efficient representation for designing and controlling large scale passive or active assembly of units (e.g. robots with complex morphologies).

Additionally, at the crossroad of Machine Learning and Evolutionary Computation, a new Reinforcement Learning approach based on modelling the user’s preferences was proposed \cite{12}, \cite{84}; in the so-called Preference-based Policy Learning, the robot demonstrates some behaviors, is informed of the user’s preferences, builds a model of the user’s preferences and self-trains to build a new behavior hopefully more satisfactory according to the conjectured user’s preferences.

Statistical Physics Perspective Basic tools from statistical physics (scaling, mean-field techniques and associated distributed algorithms, exactly-solvable models) and probability have been used to model and optimize complex systems, either standalone or combined with MABS approaches. Results are

- In the context of the ANR TRA VESTI project dealing with spatial and temporal modelling of traffic congestion we have studied in \cite{83} some specific properties of the Belief Propagation algorithm used for inference; we have proposed a way to encode dependencies between real variables with a latent binary MRF \cite{89}; we have analyzed macro-states on traffic data and how these relate to belief propagation fixed points \cite{35}.
- Also in the ANR Travesti context [36], [88] for modelling congestion at the microscopic level we have proposed a new family of queueing processes where the service rate is coupled stochastically to the number of clients. With this formulation we have been able to relate an asymmetry between acceleration and braking to some condensation mechanism. In this framework we have also proposed a large deviation formulation of the fundamental diagram of traffic flow.
- In the design of multi-objective message passing algorithms we have shown how the state-of-the-art MAXSAT solver SP-Y can be used directly by an endogeneous clause elimination procedure, to sample Pareto Front of multi-objective 3-SAT problems [87].

Multi-agent and games
- Within the InnovNation serious game project, in collaboration with Paraschool and BlueNove, we developed an e-brainstorming prototype for collaborative ideation. The prototype was tested on 150 students and is being improved to be commercialized. The realist network generator used to analyze the prototype games was also used to study the labor market, and to study the relative importance of friends and colleagues while seeking for a job without ([44] or with [64] variable information transmission speed. We have shown that friends were the most useful when the labor market was at the equilibrium (approximately the same number of jobs and applicants).
- To analyze the logs of multi-agent based simulations (for example for the InnovNation project), we developed a tool to describe homogeneous agent clusters and their evolution ([73]). We use the cluster description to build agent models and generate new simulations with this model to validate the results.
- Also in the complex social system analysis context, we applied LSA text mining tools on research projects and patent category descriptions to associate research clusters to their main research fields [70]. This was used to build a classification of the French research clusters based on their context, and especially the adequation of the cluster research specialization with the regional specialization [72].

Image understanding
- Within the context of image understanding, a new sequential recognition framework has been proposed in [10]. Sequential image understanding refers to the decision making paradigm where objects in an image are successively segmented/recognized following a predefined strategy. Such an approach generally raises some questions about the “best” segmentation sequence to follow and/or how to avoid error propagation. In [10], we propose original approaches to answer these questions in the case where the objects to segment/recognize are represented by a model describing the spatial relations between objects. The process is guided by a criterion derived from visual attention, and more precisely from a saliency map, along with some spatial information to focus the attention. This criterion is used to optimize the segmentation sequence. Spatial knowledge is also used to ensure the consistency of the results and to allow backtracking on the segmentation order if needed. The proposed approach was applied for the segmentation of internal brain structures in magnetic resonance images. The results show the relevance of the optimization criteria and the interest of the backtracking procedure to guarantee good and consistent results. From a logical standpoint, sequential object recognition is formulated as an abduction process in [14], [66]. A scene is viewed as an observation and the task of interpretation is considered as the “best” explanation considering the prior knowledge about the scene context. Towards this aim, we introduce an algebraic-based framework unifying mathematical morphology, description logics and formal concept analysis. We propose to compute the best explanations of an observation through algebraic erosion over the Concept Lattice of a background theory which is efficiently constructed using tools from Formal Concept Analysis. We show that the defined operators are sound and complete and satisfy important rationality postulates of abductive reasoning.

6.3. Crossing the Chasm

Participants: Alejandro Arbelaez, Anne Auger, Robert Busa-Fekete, Nikolaus Hansen, Balázs Kégl, Manuel Loth, Nadjib Lazaar, Marc Schoenauer, Michèle Sebag.
Due to the departure of both PhD students funded within the Microsoft-INRIA joint lab after their successful defenses (Alvaro Fialho and Alejandro Arbelaez), some of the activities of this SIG have been slightly redefined this year, with the one-month visit of Prof. Th. Runarsson (University of Iceland) in October, and the arrival in November of two new post-docs, also funded by the joint lab (Nadjib Lazaar and Manuel Loth). A new direction of research has appeared, in line with both Adaptive Operator Selection (Alvaro Fialho’s PhD) and Continuous Search (Alejandro Arbelaez’ PhD).

**Bandit-based choice of heuristics in combinatorial optimization** This new direction of research deals with heuristic choice within an existing combinatorial solver using bandit-like algorithms, and the very first results deal with scheduling problems and will be published in early 2012 [57].

**Adaptive Operator Selection** In line with his PhD work, Alvaro Fialho has successfully used his Adaptive Operator Selection method to the on-line tuning of Differential Evolution in the multi-objective case [96].

**Learn and Optimize (LaO)**, an instance-based parameter-tuning method. Though originally designed for Divide-And-Evolve framework (see Section 6.2), LaO is a generic method that learns the relationship between some instance features and the optimal parameters of the optimizer. The current version [49], [50], [51] uses Neural Network to directly learn the optimal parameters, and average performance increase compared to the default parameter set (that has won the temporal track in the IPC7 competition) is of more than 10%. On-going work uses rankSVM to learn a partial order on the features × parameter space.

**Adaptive Constraint Programming** Alejandro Arbelaez defended his PhD on May 31., led under the supervision of Youssef Hamadi and Michèle Sebag [1]. A survey of his PhD work has been published as a book chapter [74] and some of his last work was more concerned with optimizing the collaboration in distributed SAT solving in highly parallel environments [13].

**Ranking by calibrated AdaBoost** In [22], [21] we describe a learning-to-rank technique based on calibrated multi-class classification. We train a set of multi-class classifiers using AdaBoost.MH, we calibrate them using various techniques to obtain diverse class probability estimates, and, finally, we approximate the Bayes-scoring function (which optimizes the popular Information Retrieval performance measure NDCG), through mixing these estimates into an ultimate scoring function. Our method outperforms many standard ranking algorithms on the LETOR benchmark datasets, most of which are based on significantly more complex learning to rank algorithms than ours.

### 6.4. Continuous Optimization

**Participants:** Yohei Akimoto, Anne Auger, Zayed Bouzarkouna, Alexandre Chotard, Nikolaus Hansen, Ilya Loshchilov, Verena Heidrich-Meisner, Raymond Ros, Marc Schoenauer, Olivier Teytaud, Fabien Teytaud.

Our main expertise in continuous optimization is on stochastic search algorithms. We address theory, algorithm design and applications. The methods we investigate are adaptive techniques able to learn iteratively parameters of the distribution used to sample solutions. The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is nowadays one of the most powerful method for continuous optimization without derivatives. We work on different variants of the CMA-ES to improve it in various contexts as described below. In addition we have contributed to give an information geometry perspective to stochastic optimization unifying both continuous and discrete algorithms using a family of probability distribution parametrized by continuous parameters. The framework proposed in this context allow to retrieve many existing stochastic optimization algorithms when instantiated with different family of probability distributions including the CMA-ES when using gaussian distributions [85]. We have moreover clarified important design principles based on invariances [85], [11].

New algorithms based on derandomization and for mixed-integer optimization A new variant of CMA-ES to address problem with mixed-integer variables (vectors with both discrete and continuous variables) has been designed [81]. New algorithms using new selection schemes combined with
derandomization have been designed and thoroughly theoretically and empirically investigated [16], [17]. A local search algorithm using an adaptive coordinate descent has been proposed [45]. We have as well investigated how to inject solutions in CMA-ES so as to improve performances if an oracle provide good solutions [82].

**Distributed optimization** We have proposed simple modifications of evolutionary algorithms so that they reach asymptotically the optimal $\log(\lambda)$ speed-up with $\lambda$ processors and the linear speed-up $\Theta(\lambda)$ for a number of processors of order at most the dimensionality of the problem (for a pointwise solution); in particular bounding the selected population size in the Self-Adaptation algorithm and variants of this idea for fastening the decrease of the step-size in a relevant manner. All these works and more are part of Olivier Teytaud’s HDR thesis [3], defended on April 22., and also build the first part of Fabien Teytaud’s PhD [2], defended on December 8.. We wrote a chapter on lower bounds for distributed derivative free optimization in [79].

**Benchmarking** We have continued our effort for improving standards in benchmarking and pursued the development of the COCO - COmparing Continuous Optimizers platform [80]. We are organizing for the GECCO 2012 conference the Black-Box-Optimization Workshop⁴.

**Optimization with meta-models and surrogate** We have investigated optimization using a coupling of CMA-ES and surrogates and applied it for the optimization of well placement [6]. We have proposed a new meta-model CMA-ES for the optimization of partially separable functions [19] and shown that it improves performances for solving the well placement problem [19].

**Hyperparameter optimization** In [18] we present hyper-parameter optimization results on tasks of training neural networks and deep belief networks (DBNs). We optimize hyper-parameters using random search and two new greedy sequential methods based on the expected improvement criterion. The sequential algorithms are applied to the most difficult DBN learning problems and find significantly better results than the best previously reported.

**Multi-objective optimization** We have investigated theoretically multi-objective algorithms based on the hypervolume [4] and proposed new selection operators based on tournament and multi-armed bandit framework [46].

**Multimodal optimization** We have shown in [58] a simple algorithm (a $(1+1)$-ES with quasi-random restart and murder operator), which, at least in the sequential case, performs as efficiently as much more tricky algorithms.

**Mathematical bounds for noisy optimization** The paper [27] shows upper and lower confidence bounds and/or experiment algorithms in the noisy optimization setting; in particular we compared an optimization algorithm based on bandits and an surrogate-model version; whereas the bandit approach is much faster if the noise decreases quickly to zero around the optimum, the surrogate-model version is faster if the noise does not decrease to zero.

### 6.5. Optimal Decision Making

**Participants:** Olivier Teytaud, David Auger, Michèle Sebag, Cyril Furtlehner, Jean-Baptiste Hoock, Nataliya Sokolovska, Fabien Teytaud, Hassen Doghmen, Jean-Joseph Christophe, Jérémie Decock.

- Monte-Carlo Tree Search (MCTS) and Upper Confidence Trees (UCT) are main areas of the team. In particular, we ultra-weakly solved 7x7 Go by winning 20 games out of 20 against professional players in 7x7 Go, thanks to a Meta-Monte-Carlo-Tree Search [24]. The wins were with komi 9.5 as white, and 8.5 as black, suggesting that the ideal komi in 7x7 is 9. We also applied this algorithm to the recent “NoGo” framework, aimed at challenging MCTS for a game which looks like Go but with very different goals; our paper [26] was the first one applying MCTS to NoGo and now all strong programs use the MCTS approach for NoGo. We extended RAVE (Rapid Action Value Estimates) to the continuous settings [29]. In his PhD [2], Fabien Teytaud proposed several generic

---

⁴ see http://coco.gforge.inria.fr/doku.php?id=bbob-2012
improvements of MCTS, including the use of (fast) decisive and anti-decisive moves for games, and applied it to the game of Havannah. An industrial application (to energy management) is proposed in [71]. A MCTS version for partially observable problems with bounded horizon was proposed in [86]. This version is proposed for the two-player case, but for simulations starting at the root; a version in the one-player case, starting from an arbitrary state (and therefore much more efficient for large horizon) is proposed in [30]. This work is extended by a belief state estimation by constraint satisfaction problems in [62]. Other developments and research around MCTS/UCT are described in the MoGo module.

- A related important algorithm is Nested Monte-Carlo; we got state of the art results for some traveling salesman variants with a very simple algorithm in [56].
- Fundamental analysis of partially observable games: we proved in [5] that partially observable games are undecidable (result also presented in the BIRS 2010 workshop and the Bielefeld seminar on Search Methodologies), even in the case of finite state spaces and deterministic transitions. This unexpected result is a priori a contradiction with known decidability results; this emphasizes the subtle difference between the classical decision problem (the existence of a strategy winning certainly, whatever may be the strategy of the opponent), which is used is most analysis, and the choice of the move with optimal winning probability. We pointed out that the relevant decision problem is, with no doubt, the latter; that the other decision problem has just been used because it is equivalent to choosing optimal play in the case of fully observable games; and, most importantly, that partially observable games are in fact undecidable, even in the finite deterministic case. On the other hand, on restricted settings, we have shown by some simple lemmas lower and upper bounds on the value of some partially observable games [63]. We extended Monte-Carlo Tree Search to the case of short-term partial information in [61]; this was successfully applied to the Urban Rival game, a widely played internet card game (now 17 millions of registered users) from a French company.
- Tuning of strategies: tuning strategies is a noisy optimization problem in which the convenient “variance of noise decreasing to zero around the optimum” usually does not hold. We have shown that in such a setting, the local bandit-style algorithms are slower than surrogate models; this is detailed in the continuous optimization part.
- We organized various computer-Go events, as due to the fame of our program MoGo we are often invited for such events; reports can be found in [95].
- We developed the “double progressive widening” trick, which is aimed at making consistent an algorithm from the finite case to the continuous stochastic case; we got good results in [60] on Q-Learning (with no mathematical proof) and on MCTS [28] (mathematical proof to be submitted soon).
- We have also worked on Nash equilibria of Matrix Games, where we proposed an algorithm for finding Nash equilibria faster when the Nash equilibrium is sparse [48], [47]; a mathematical proof is ready and will be submitted soon.
- Some works are in progress around applications of previous tools to active learning; active learning has also been investigated through conditional random fields in [59].
- Another related work, with motivations from autonomous robotics, combines the exploration of the search space through UCT, with an explicit model of the safe regions explored so far, called Deja-Vu. The Deja-Vu is used to constrain the exploration, mostly in the random phase, and is updated from the current explorations [67].
- The Ilab “Metis” just started; it’s an Ilab between Tao, the Inria-Saclay team MaxPlus, and the SME Artelys http://www.artelys.com for a joint work on numerical libraries in Energy Management.

6.6. Large and Deep Networks

Participants: Ludovic Arnold, Sylvain Chevallier, Anthony Mouraud, Hélène Paugam-Moisy, Sébastien Rebecchi, Michèle Sebag.
Deep networks: RBM or AA  The two main families of deep networks are implemented and studied by TAO: stacked RBM (Restricted Boltzmann Machines) and stacked AA (Auto-Associators).

Learning sparse features for deep networks  Inspired by the theory of compressed sensing and beyond the common methods based on dictionary learning, we have proposed to learn sparsity and accuracy simultaneously by alternating two constraints on the weights of an Auto-Associator [55].

Spiking neuron networks for swarm robotics  The model "SpikeAnts" [91] has been applied to a spatial robotic environment [23], in collaboration with Nicolas Bredeche (see Section 6.2), and has demonstrated even more its interest in the context of swarm robotics.
TYPICAL Project-Team

6. New Results

6.1. Semantics of the Calculus of Inductive Constructions

Participant: Bruno Barras [Contact].

Bruno Barras has formalized the meta-theoretical study of strictly positive inductive types. This was built upon the previous work on specific instances: natural numbers and Brouwer ordinals. The main idea of the model construction is to use the property that every strictly positive inductive definition can be encoded in the parameterized type of trees (the so-called W-types). Such tree-types can themselves be encoded as partial functions from paths to labels. The soundness of this translation gives a way to build the closure ordinal of any strictly positive inductive definition.

Bruno Barras has then modelled the inductive families (also called inductive types with indices). He has been able to prove formally the previously known result that inductive family can be constructed in two steps: first build a carrier type (inductively) which is oblivious of the indices, and then define each member of the family as a subset of the carrier type by enforcing the constraints generated by the indices.

He also started to investigate advanced features of inductive definitions, like the possibility to have non-uniform parameters. When this feature was introduced in Coq, it was thought as a conservative one, but the formal analysis showed that this was not obvious. The consistency model could be extended (with one auxiliary result not yet encoded formally). This shows that non-uniform parameters do not extend much the expressivity of Coq, but the strict equivalence remained as an open problem.

6.2. Relative Strengths of set theory and type theory

Participants: Bruno Barras [Contact], Benjamin Werner.

Bruno Barras also formalized common translations in proof theory: negated translations and Friedman’s A-translation. This was used to build a model of (classical) ZF set theory in Coq extended with one axiom called TTDA (Type-Theoretical Description Axiom). This was done in two steps: first build a model of $IZF_C$ (ZF with the collection axiom but not the excluded-middle) in Coq extended with TTDA, and then encode ZF in $IZF_C$, as shown by Friedman.

The converse result: an interpretation of Coq +TTDA in ZF (with one inaccessible cardinal!) seems not possible, as TTDA in a classical setting gives a (weaker) form of the axiom of choice. Bruno Barras as devised a new axiom (called the Type-Theoretical Collection Axiom) that still allow the ZF interpretation above, but he hopes that its consistency can be proved in ZF extended with one inaccessible cardinal.

Benjamin Werner has worked with Gyesik Lee on set-theoretical models of Coq’s type theory. This work is described in a paper published in the LMCS journal [18].

6.3. A Consistency model of Coq extended with decision procedures

Participants: Bruno Barras [Contact], Qian Wang.

Bruno Barras and Qian Wang are working on the construction of a model for the Calculus of Constructions extended with the type of natural numbers. The definitional equality has been extended to include all equations derivable in Presburger arithmetic. Compared to previous work, this model can support strong eliminations. Since strong eliminations and extensions of the definitional equality with non-satisfiable equations (for instance $0 = 1$) leads to non-normalizing terms, it was necessary to give a precise account of Presburger arithmetic, seen as a specific instance of first-order logic. This work is described in a paper published in the proceedings of the LICS conference [21].
6.4. Towards a concurrent architecture for the Coq kernel

**Participants:** Bruno Barras [Contact], Enrico Tassi.

In the context of the Paral-ITP ANR project, Bruno Barras and Enrico Tassi have started to implement a kernel of Coq where the process of constructing and checking the proof of a lemma can be executed in a parallel thread.

6.5. Physics of computation

**Participant:** Gilles Dowek.

Together with Pablo Arrighi, Gilles Dowek has extended Gandy’s theorem to quantum physics, by giving a new definition of the notion of finite density of information in this setting. This work has been presented at the congress QIPC [13].

6.6. Binders

**Participant:** Gilles Dowek.

Together with Jamie Gabbay, Gilles Dowek has given a translation of permissive nominal logic to Higher-order logic and proved its soundness and completeness. This work is described in a paper published in the Transactions on Computational Logic [15].

6.7. Interfacing Coq with SMT solvers

**Participants:** Germain Faure, Chantal Keller [Contact], Assia Mahboubi, Benjamin Werner.

This is work in close collaboration with the Marelle team (INRIA Sophia Antipolis). The starting point of this work is to note that SMT solvers, deciding the Satisfiability Modulo Theories, are in constant evolution to take into account new decision procedures as well as theories. These systems are rather complex and it is now clearly established that they all contain bugs. The standard approach is to ask the SMT solver to append to the decision result a certificate that can be checked by another tool.

In this context, we are using Coq to check the certificate. The approach is based on computational reflection. The checker is written in Coq, and its architecture is modular and extensible.

We are now able to check certificates coming from the ZChaff SAT solver and from the veriT SMT solver developed at INRIA Nancy – Grand - Est. Proofs established by the SMT tool for the theories of congruence closure and linear arithmetic are checked in short time, overtaking the state of the art in terms of time performance. We also use certificates to build a new Coq tactic that can safely call an external SMT solver, thus increasing Coq’s automation. This tactic is new since it is a decision procedure that combines both linear integer arithmetic and equality of uninterpreted functions. This work is described in a paper published in the proceedings of the CPP2011 conference [25].

6.8. SMT techniques for optimization problems

**Participant:** Germain Faure [Contact].

The TypiCal team has collaborated with the sysmo team at the Laboratoire d’Informatique de l’École Polytechnique in order to integrate the use of automated tools like SMT solvers in the resolution of optimization problems. The case study was a problem of large scale energy management with various constraints, proposed at the ROADEF 2010 challenge won by the sysmo team. We investigated how to delegate to the SMT tool part of the resolution of constraints. A first conclusion of this experiment is that solving optimization problems represents a more important part of the computation time than first expected. As SMT solvers are not geared toward this class of problems, their performance were not satisfactory. This nonetheless opens new perspectives for the development of SMT tools in order to adapt their internal decision procedures to this new kind of benchmarks. We consider that significant progress in that direction could be easily obtained.
6.9. Formal correctness of embedded programs  
**Participants:** Gilles Dowek [Contact], Pierre Néron.

Pierre Néron is working on program transformations that remove the operations which create most of the approximations during the computation on floating point numbers, namely square roots and divisions. This kind of formal tool aims at increasing the confidence in embedded programs. The idea of this transformation comes from the elimination of the quantifier on real closed fields, hence the first task is to define a minimal but useful language on which the transformation will apply and then to extend this transformation on formulas to this whole language. Keeping the size of the code produced by this transformation in an acceptable range was a challenging issue in this work. The next objective is to write a formal proof ensuring that the transformation is correct. This work will be done in collaboration with the NASA Langley research center in the Formal Method team: Pierre Néron will visit this center for one month in January 2012.

This work is described in a submitted paper [30].

6.10. Formal proofs for convex optimization problems  
**Participants:** Benjamin Werner [Contact], Victor Magron.

Victor Magron is working on the integration of tools that can deal with inequalities on semi-symbolic expressions with real numbers inside proof assistants like Coq.

In particular, he is working on new means to provide formally established bounds for multivariate inequalities, using methods inspired from the convex optimization literature like sums of squares (SOS) and the related semi-definite programming (SDP) relaxation.

He has implemented in OCaml a new algorithm which detects and computes automatically the possible bounds of a given expression. He has tested the approach using benchmarks largely built from inequalities issued from the formal proof of Kepler conjecture (by Thomas Hales). The algorithm computes approximation of transcendental functions by solving sum of squares problems, delegated to an external, dedicated tool. The next step of this project is to certify the correctness of these computations using the Coq system.

He has also improved a Coq tactic based on the external computation of decompositions into sums of squares originally developed by Fréderic Besson (INRIA Rennes - Bretagne Atlantique). The improvement consists in linking this tactic with a tool developed by David Monniaux (Verimag).

6.11. Formal proof in real algebraic geometry  
**Participants:** Assia Mahboubi [Contact], Cyril Cohen.

Cyril Cohen and Assia Mahboubi have completed the first formal proof of quantifier elimination for the theory of real closed fields. This work includes a significant part of infrastructure code for ordered algebraic theories, intervals, and polynomials. This work is described in a submitted paper [29].

Cyril Cohen has implemented in Coq a construction of real algebraic numbers and proved it had a structure of discrete Archimedian real closed field, in the sense of the previous proof of quantifier elimination. Beside the computational interest of real algebraic numbers, this construction both legitimates the abstraction chosen for the proof of quantifier elimination and provides a basis for complex algebraic numbers needed for the completion of the formal proof of the Feit-Thompson theorem. This work is described in a paper to appear in the proceedings of the JFLA2011 conference.

6.12. Constructive mathematics  
**Participant:** Cyril Cohen [Contact].

In collaboration with Thierry Coquand, Cyril Cohen has come up with a constructive proof of a generalization of the fundamental theorem of Algebra. This work show how to formalize the algebraic closure of an arbitrary real closed field. In particular, it can serve as a basis for the construction of complex algebraic numbers from the real algebraic numbers. This work is described in a submitted paper [28].
6.13. Intersection types
Participants: Alexis Bernadet [Contact], Stéphane Lengrand [(CNRS, Lix)].

Alexis Bernadet and Stéphane Lengrand have studied a typing system for the \( \lambda \)-calculus with non-idempotent intersection types. As it is the case in (some) systems with idempotent intersections, a \( \lambda \)-term is typable if and only if it is strongly normalizing. Non-idempotency brings some further information into typing trees, such as a bound on the longest \( \beta \)-reduction sequence reducing a term to its normal form. These results are presented in Klop’s extension of \( \lambda \)-calculus, where the bound that is read in the typing tree of a term is refined into an exact measure of the longest reduction sequence. This complexity result is, for longest reduction sequences, the counterpart of de Carvalho’s result for linear head-reduction sequences. This work is described in a paper published in the proceedings of the FOSSACS 2011 conference [22].

Alexis Bernadet and Stéphane Lengrand have also revisited models of typed \( \lambda \)-calculus based on filters of intersection types. By using non-idempotent intersections, they simplify a methodology that produces modular proofs of strong normalization based on filter models. Non-idempotent intersections provide a decreasing measure proving a key termination property, simpler than the reducibility techniques used with idempotent intersections. Such filter models are shown to be captured by orthogonality techniques: we formalize an abstract notion of orthogonality model inspired by classical realizability, and express a filter model as one of its instances, along with two term-models (one of which captures a now common technique for strong normalization). Applying the above range of model constructions to Curry-style System F describes at different levels of detail how the infinite polymorphism of System F can systematically be reduced to the finite polymorphism of intersection types. This work is described in a paper published in the proceedings of the CSL 2011 conference [23].

Participants: Alexis Bernadet [Contact], Stéphane Lengrand [(CNRS, Lix)].

We introduce here an alternative definition of Hyland’s effective topos, based on a realizability framework with two levels of abstraction: a low level and a high level. With this definition, the proof that this framework forms a topos is almost as simple as proving that the category of sets is a topos. Moreover, the high level of the framework can be directly used as a model of higher-order intuitionistic systems. We can then craft a programming language based on topos theory, which can be given a constructive semantics. In such a programming language, we can only write functions that terminate, as in proof assistants like Coq, so the language cannot be Turing-complete. The main advantage of having a programming language based on topos theory over more usual intuitionistic systems such as Martin-Löf type theory is the notion of equality: it is extensional, has proof-irrelevance, and allows the axiom of unique choice.

This work has been presented at the Chocola-Ens Lyon seminar in December 2011.

6.15. A formal proof of the Feit-Thompson theorem
Participant: Assia Mahboubi [Contact].

Assia Mahboubi has pursued her work in the Mathematical Component team lead by Georges Gonthier at the Microsoft Inria Joint Centre. She has finished the formalization of the Wielandt fixpoint theorem, which is one of the key results at the interface between the two components (local analysis and character theory) of the published revised proof of the Feit-Thompson theorem. The proof of the Wielandt theorem was difficult to formalize because it requires a challenging combination of advanced theories with sophisticated constructive formalization: group representation, module theory, linear algebra and characters.

The documentation of this formalization, as well as the current state of the whole formal proof can be found on the webpage of the Mathematical Components project.

6.16. A formal library for polynomial arithmetics
Participant: Assia Mahboubi [Contact].
Assia Mahboubi has worked on a modular formal library devoted to the divisibility theory of polynomials. The aim of this library is to provide a solid basis for further formal developments involving algorithms on polynomials, in particular to cover the cases when the coefficients of the polynomials involved are equipped with a structure weaker than the structure of field required by the standard Euclidean algorithm.

The documentation of this formalization can be found on the webpage of the Mathematical Components project.

6.17. Weak Memory Models

Participant: Assia Mahboubi [Contact].

Assia Mahboubi has collaborated with Jade Alglave (Oxford University) and has programmed a complete formalization in Coq of the semantic proposed by Jade Alglave in a PhD for weak memory models. This work is described in [27].